

Linked Data 技術による異常状態オントロジーと 生命医学系リソースの統合への試み

A Linked Data Approach for Integration of Ontology of Abnormal States with Biomedical Resources

山縣 友紀¹ 古崎 晃司¹ 溝口 理一郎² 今井 健³ 大江 和彦⁴ 駒谷 和範¹

Yuki YAMAGATA¹, Kouji KOZAKI¹, Riichiro MIZOGUCHI², Takeshi IMAI³,
Kazuhiko OHE⁴, and Kazunori KOMATANI¹

¹ 大阪大学産業科学研究所

¹The Institute of Scientific and Industrial Research (ISIR), Osaka University

² 北陸先端科学技術大学院大学

² Advanced Institute of Science and Technology

³ 東京大学医学部附属病院 企画情報運営部

²Department of Planning, Information and Management, The University of Tokyo Hospital

⁴ 東京大学大学院医学系研究科

⁴Graduate School of Medicine, The University of Tokyo

Abstract: Recently, biomedical terminologies and ontologies have been publishing a wide variety of data. For the integration of them, it is important to overcome the conceptual discrepancy. In this article, we discuss ontological issues of integrating knowledge of abnormal states in existing biomedical resources. Next, we show an approach for the solution with our ontology of abnormal states. Then, by applying Linked Data technique, we develop a prototype system for linking our ontology as a hub of existing resources. Our approach will bring benefits for promoting interoperability across heterogeneous data according to the meaning of concepts and contribute to developing knowledge integration infrastructure for disease.

1 はじめに

近年, Linked (Open) Data 技術を用いて公開されるデータ数はますます増加しており, 生命科学ドメインでは, 特にその動きは活発である[1]. また, 臨床ドメインにおいても電子カルテシステムの導入など, 情報のシステム化が急速に進められるとともに, 組織を横断的した情報共有と相互運用性の確保が不可欠となってきた。

筆者らは, これまで臨床医学オントロジープロジェクトにおいて, オントロジー工学理論を基礎として[2], 疾患オントロジーの開発を行ってきた[3,4]. その一環として, 疾患定義における異常状態に着目し, 異常状態オントロジーの構築を進めてきた[5]. 今後, 既存のオントロジーや用語集の有用なデータを積極的に活用し, 異常状態知識を統合することは疾患解明に向けた知識基盤の構築において重要とな

る.

そこで本研究では, まず, オントロジー工学的な観点から統合上の課題について議論するとともに, 異常状態オントロジーを用いたアプローチについて提案する. 次に, 典型的な疾患を事例とした異常状態知識の統合の試みを報告し, 更に, 統合実現に向けた異常状態オントロジー-Linked Data 化とその有用性について考察する.

2 統合における課題とアプローチ

生命医学分野では, 様々な用語集やオントロジーが開発されている. それらのうち, 異常状態概念を含むものとして, オントロジーでは, PATO (Phenotypic Quality Ontology) [6], HPO (Human Phenotype Ontology) [7], 用語集では SNOMED-CT¹

¹ <http://www.ihtsdo.org/snomed-ct>

(Systematized Nomenclature of Medicine -Clinical Terms)と MeSH(Medical Subject Headings)²が存在する[5]. そこで、これら4つの既存のリソースを対象に、異常状態に関する知識を統合するための課題について、オントロジー工学の観点から考察する。

2.1 統合における課題

2.1.1 概念レベルの不一致

専門家は、通常、専門領域に特化した用語集やオントロジーを構築する。そのため、自身の構築した概念の粒度や抽象レベルが外部リソースのものとは一致するとは限らず、そのままでは関連づける事が難しいという問題が生じる場合がある。

例えば、PATO、HPO はともに表現型オントロジーとして知られている。表現型とは、生物のもつ性質のうち、実験などを通じて観測しうるものをいう。PATO は生物種横断的に参照できる概念を提供することを目的としており、「部品の欠損」、「濃度減少」のような汎用的な概念を体系化している。そのため、ある生物種の特定の器官で起こる概念は含まれていない。一方 HPO はヒトに特化し、人体の各器官で観察される異常(例:冠動脈閉塞、肺高血圧)を提供している。しかし、それらを一般化した概念は持っていないため、PATO で定義されているような概念レベルの大きく異なるものをそのまま関連付けることは困難である。

ここで、大規模数の概念をカバーする事で知られる臨床用語集 SNOMED-CT や MeSH を用いれば、これらのレベルをうまくつなぎ合わせる事ができるのではないかと考えられる。しかし、これらの用語集は、もともと病理学会や国立図書館が開発を行ってきたため、オントロジー理論に準拠しておらず、上位オントロジーももたない。そのため、is-a 関係自体のエラーや、多様な視点の下、多重継承を認めた分類階層となっている。その結果、概念間の関係が非常に複雑化し、必要な概念を同定することさえ困難である。したがって、これらの用語集に概念レベルのギャップを埋める役割は期待できない。

2.1.2 特性と属性の不分離

次に、PATO、HPO は上位オントロジーとして BFO (Basic Formal Ontology) [8]を採用している。BFO では、哲学者による存在論としての性質を重視するため、特性(以下、P で表す)と属性(以下、A で表す)がオントロジー工学的な観点において識別されていない。PATO もそれに従い構築されているため、例えば、「面積縮小 (P)」という性質を表す特性と、

物理的なパラメーターに対応する「面積 (A)」のような属性の区別がされていないことがわかった[5]。その結果、PATO の is-a 階層では上位概念が複数存在し、このような多重継承は、PATO と他のリソースと統合する際に問題が生じる可能性がある。

例えば、HPO の「冠動脈狭窄」という用語は、冠動脈の面積が縮小下状態を指す。そこで、PATO の「面積縮小」を上位概念との関連付けが適切と考えられる。ここで、PATO 「面積縮小」は、属性「面積」と特性「サイズ縮小」の2つを上位概念として持つ。そのため計算機で自動的に統合処理すれば、さらに面積縮小の上位概念として「面積」も関連づけられ、その結果「冠動脈狭窄 is-a 面積」のような、異常状態としては間違った is-a 関係が推論されてしまう可能性がある。異常状態は変化しうるため、どのような観点(A)で、どのような値(V)をもつかという点において、「面積縮小 is-a 面積」ではなく、「面積縮小 is-a サイズ縮小」を本質として捉えるべきである。

また、PATO や HPO では、特性に対応する属性と値に分離するような表現形式をもたないため、例えば「狭窄」の場合、「断面積」(A)とその値「小さい」(V)に分けることができない。検査データや実験データとの統合を考慮するならば、パラメーター(属性)とその値(V)の分離は科学的な客観性、透明性の確保の上で必須といえる。

2.1.3 概念の不明確性: 疾患と異常状態の混在

外部データと統合する場合、概念自体が曖昧で不明確であると、本来は概念的に区別されるべきものが、統合の際、システム上では同一と判断される恐れがある。

臨床ドメインでは、異常状態と疾患の識別は、電子カルテシステムにおける診療記録において特に重要となる。しかし、HPO では定義内容が記載されていないものも多い。また、MeSH では、例えば「心筋虚血 Myocardial Ischemia」の場合、定義では「心筋の血流不足から生じる心機能の異常」というように異常状態とされているが、同義語に「虚血性心疾患 Ischemic Heart Disease」という疾患が入っているというように、疾患か異常状態か不明なものが多い。

さらに、MeSH、HPO および SNOMED-CT はいずれも疾患と異常状態が混在した階層化が行われている。

SNOMED-CT は、近年国際的な標準医療用語として各国の病院間の電子医療情報システム等に使用されつつある。しかし、SNOMED-CT の分類階層では、異常状態、疾患に加え、「所見」など複数の上位概念をもつことを認めている。概念が疾患か異常状態か不明なまま使用されると、記録自体に信頼性が懸念

² <http://www.nlm.nih.gov/mesh/meshhome.html>

される。

表現型のオントロジーである HPO の場合においても同様に、表現型とは言えない疾患概念が混在している。例えば、「冠動脈狭窄」の上位概念は、「冠動脈疾患」となって、さらにその上位には異常状態や疾患が混在する。

このように、リソースを統合するには、疾患と異常状態というような概念の混在を避けることが必要であり、臨床の場での利用に耐えうるものとするには、その違いについて明示する必要がある。

2.2 異常状態オントロジーによるアプローチ

本章では、著者らが構築中の異常状態オントロジーの概要を紹介するとともに、本オントロジーを用いた生命科学系リソースとの統合に向けた課題解決のためのアプローチを提案する。

2.2.1 異常状態オントロジー概要

異常状態オントロジーは、ドメインに依存しない汎用的な概念から、人体各器官固有の概念、さらに臨床ドメインにおける各診療科の専門医が疾患を定義する上で必要とする異常状態まで、3層モデルとして組織化を行っている[5]。

レベル1：汎用異常状態（対象物非依存異常状態）

レベル2：対象物に応じて共通に現れうる異常状態

レベル3：疾患を構成する異常状態

最上位（レベル1）の汎用異常状態は、特定のドメインに依存しない多くの対象物に見られる（対象物非依存）共通の異常状態を定義している。汎用異常状態は、構造関連異常、機能関連異常、パラメータ異常などに大別される。概念例として、面積縮小、変形、機能低下、圧力増加等が挙げられ、これらは生物医学分野以外でも用いられる一般的な基本概念といえる。

本異常状態オントロジーのレベル1は、PATOの汎用的な概念と対応すると考えられる。そこで、実際にPATOの対応概念へ関連付けることとした。

次に、レベル2の対象物に応じて共通に現れうる異常状態では、異常の対象となるオブジェクト（対象物）に固有の異常状態が定義されている。レベル2のis-a階層は、レベル1で定義された汎用異常状態からその定義内容を継承しつつ、対象物にしたがって特殊化することで定義される。レベル2の上位では、管状構造物の異常や弁構造物の異常など、複数ドメインに共通する汎用的な構造物を対象とした異常状態が定義される。例えば、管状構造物では、管の断面積が縮小する「狭窄」という異常が起こりやすいという共通の性質をもつといえる。そこで、

レベル1の汎用異常状態である「面積縮小」を特殊化することでレベル2では「管狭窄」を定義することとした。レベル2の下位では、専門ドメインの対象物において現れる共通の異常状態が定義されている。本研究では、医療ドメインにおける人体解剖構造物を対象としている。例えば、管状構造物の異常の場合、「管狭窄」を特殊化した人体固有の「血管狭窄」が定義され、管の断面積が縮小する性質を継承しつつ、各診療科の専門医が必要とする粒度まで特殊化を行い、is-a階層を構築している。

HPOは人体の表現型を対象としているので、本異常状態オントロジーのレベル2の下位に位置する人体の各器官依存の概念と対応するといえる。そこで、本レベル2下位の異常状態概念をHPOの対応する概念に関連付けることとした。MeSH, SNOMED-CTについても同様の理由から、本レベルと関連付けることとした。

さらに、レベル3では、疾患を構成する異常状態を定義しており、例えば、「虚血性心疾患における心筋虚血」、「糖尿病における高血糖」のように、疾患コンテキストに依存して定義されている。そこで、これらと対応する概念が外部リソースで存在する場合、関連付けることとした。なお、レベル3は、疾患そのものとは明確に区別している。疾患は、疾患オントロジーで臨床の専門医が中心となり定義を行っている[2]。そこで、外部リソースについては、疾患概念と定義から判断できるものは、疾患オントロジーとマッピングすることとし、疾患を構成する異常状態と認定できる概念については、異常状態オントロジーのレベル3と関連付けることとした。

2.2.2 異常状態の統一表現

異常状態オントロジーでは人体に現れる多様な異常状態に対し、一貫性を保持するための統一表現記述枠組みとして、特性属性分解によるモデルを提案している[9]。本表現モデルでは、異常状態をまず特性（Property, (P)）と特性値（Property Value (Vp)）で捉え、さらに特性を対応する属性（Attribute (A)）とその値（属性値）（Attribute Value (V)）に分解する。属性値は、定量値（Quantitative Value (Vqt)）、定性値（Qualitative Value (Vql)）またはカテゴリ値（Vcat）のいずれかをとることで特性表現、定性表現、定量表現に対応が可能である。

異常状態表現モデルでは、基本表現として、対象となるオブジェクト（Object (O)）と組み合わせたOAVトリプル表現〈対象物 (O), 属性 (A), 属性値 (V)〉を用いており、さらに必要に応じて注目物に関する情報を「副対象物（Sub Object (SO)）」を導入した拡張表現によって、比などの複雑な表現に

についても対応可能とし、柔軟かつ一貫性を保持した表現となっている。

本表現モデルは、検査データのような定量表現から疾患定義における異常状態を表現するための特性表現まで、臨床上の表現すべてについて計算機で処理が可能である。さらに、本研究では、特性と属性を区別した表現モデルを採用しているため、性質として下位概念がその本質属性を継承しているかどうかも考慮したうえで外部リソースと関連付けることが可能になる。

3 統合の試行

2.2 節で示したアプローチが実際に統合の解決手段となりえるかを検討するため、具体例を用いて、異常状態オントロジーのレベル1からレベル3まで体系化された異常状態と、疾患オントロジーおよび外部リソースとの統合を試みた。対象は、専門医との協議の下、3 診療科（循環器、神経内科、消化器内科）の代表的な12疾患とそれを構成する異常状態を選定した。

異常状態オントロジーと外部リソースとの関連付けは、表層上ラベルの文字列完全一致で機械処理した後、その結果を2章で議論したオントロジー工学的な観点に基づいて精査し、属性の継承関係を考慮しながらマッピング内容を修正することで行った。

3.1 概念レベル別統計

異常状態オントロジーの各レベル（レベル1-3）における外部リソース（PATO, HPO, MeSH, SNOMED-CT）との統合結果を表1(a)に示す。

異常状態オントロジーのレベル1汎用異常状態134概念のうち52概念について、PATOと1対1対応で関連付けられた。また、レベル2対象物に応じて共通に現れうる異常状態では、レベル2上位の2

表1 異常状態に関連する既存リソースとのマッピングに関する統計

(a) 異常状態オントロジーと外部リソースとの概念レベル別マッピング数

概念レベル	異常状態オントロジー	PATO	HPO	MeSH	SNOMED-CT
レベル1	134	52	0	2	0
レベル2	145	2	27	28	17
レベル3	107	0	0	0	0
総計	386	54	27	30	17

(b) HPO および異常状態オントロジーから PATO へのマッピング内訳

	HPO		異常状態オントロジー	
	個数	割合	個数	割合
PATOへマッピングされていたもの	14	52%	20	74%
マッピングされないもの	13	48%	7	26%
合計	27	100%	27	100%
マッピングされないもの内訳				
対応概念がPATOに存在せず	7	26%	7	26%
対応概念がPATOに存在	6	22%	0	0%

概念と PATO の用語が対応することがわかった。これは、細胞壊死、組織炎症という全生物種に共通の器官で表れる汎用的な性質であった。また、PATOはレベル2の下位にある人体の器官依存の異常状態に関する概念やレベル3疾患を構成する異常状態とはマッピングされないことが確認された。

次に、レベル2の人体器官依存の異常状態はHPO, SNOMED-CTと対応し、HPO, SNOMED-CTはレベル1の汎用的な概念とは直接対応が付けられないことが確認された。MeSHもほぼ同様の結果であった。

ここで、異常状態オントロジーでは、2.2節で述べたように、レベル2上位に管状構造物や弁構造物のような汎用構造物依存の異常状態を定義している。本概念は、PATO汎用レベルの概念と、HPOの人体の器官固有の概念とのギャップを埋める役割を果たすことがわかった。一例を挙げると、HPO「動脈狭窄」は、異常状態オントロジーとはレベル2下位の「動脈狭窄」とマッピングされ、動脈の断面積が縮小した状態を意味する。一方、PATOの「面積縮小」はレベル1の「面積縮小」とマッピングされる。ここで、異常状態オントロジーではレベル2上位「管狭窄」を設けており、両者の中間層として医療ドメイン固有の知識を複数ドメインに共通する汎用的な知識まで円滑につなぐこと役割を果たしているといえる(図1①)。さらに、本概念は「大腸狭窄」など人体の様々な管状構造物の異常との関連性を見出す役割としても機能する。

3.2 外部リソース間のマッピングに関する評価および考察

HPOは近年、各種オントロジーとの相互運用を試みている[7]。そこで、前節で示した異常状態オントロジーと対応がついたHPOの27概念のうち、HPOから直接PATOへマッピングされているものがあるか調査した。その結果、27概念中14概念(52%)について直接マッピングされていることがわかった(表1(b))。

次に、PATOへのマッピングがされていないHPOの13概念について精査すると、13概念中6概念について、PATO内に、本来ならば上位概念としてマッピングできる概念が存在することがわかった。例えばHPO「高血圧 Hypertension」の場合、それに対応する「圧力増加 increased pressure」という汎用的な概念がPATOには存在し、HPO「発熱 Fever」では、対応する概念として「温度上昇 increased temperature」がPATOに定義されている。しかし、これらは表層上のラベルが一致しないため、概念レベルの違いを超えてHPOからPATOへ直接マッピング処理する事ができずに残っていると推察される。

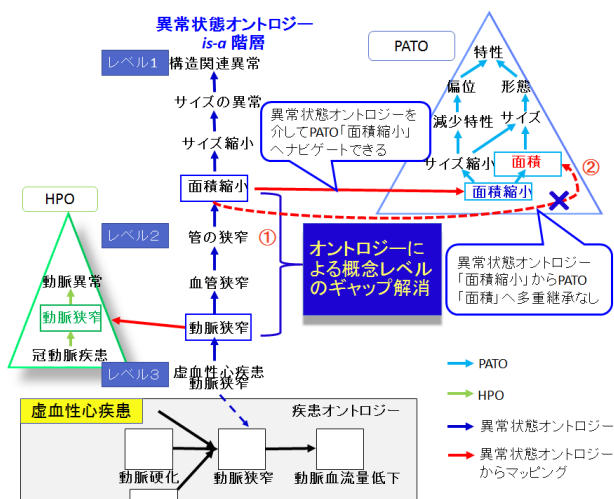


図1 表現型オントロジーとの概念レベル別統合の例

一方、異常状態オントロジーでは、これら6概念についてはすべて PATO へマッピングしており、本異常状態オントロジーを介した場合、HPO から PATO へのマッピング数は20概念(74%)となり、マッピング率が向上することがわかった。なお、本異常状態を介してもHPOからPATOにマッピングされない残りの概念については、胸痛や吐き気など、観測できない感覚異常に関する概念であり、PATOにはそのような概念に対応する汎用概念は提供されていないことが原因である。

このように、異常状態オントロジーは、すでに関連付けられたリソース間の互いの立ち位置を客観的に示すとともに、適切な概念にナビゲートする役割を果たすことがわかった。

3.3 概念の明確化による信頼性の向上

本異常状態オントロジーは特性を属性分解する表現形式をとっているため、例えば「狭窄」のような臨床用語が「断面積(A)、小さい(V)」と属性とその値に分解されることで、概念の透明性が増すとともに、どのような汎用異常状態を上位概念としてもち、それを特殊化したのが、外部ユーザーにとっても客観的に把握可能である。今回の試行では、特性と属性の分離により、本異常状態と他のリソースを関連付けたとしても、PATOのように「狭窄 is-a 面積縮小 is-a 面積」というような推論で異常状態としての観点においてエラーが起こる心配はないことを確認した(図1②)。

Linked Data では、任意のデータの関連付けを許しているが、不明瞭な概念では再利用が困難となる。概念レベルと概念の捉え方を明示でき、それをユーザーが参照、選択できるような枠組みがあれば、なぜ関連付けられているかが解釈可能となり、信頼性が生まれる。そのような対象の理解のための視点を

与え、互いの関連性付与を支える骨格がオントロジーであり、安心して実践で使えるための妥当性を担保する。

3.4 試行結果のまとめ

試行実験では対象とした数が少数ではあるが、異常状態オントロジーは、既存リソースに対し、1)概念レベルの付与、2)レベル横断的に適切な概念へのナビゲート、3)関連づけの根拠となる本質属性の明示を提供できる事を確認した。本手法はスケールアップしても適用可能であり、知識統合基盤づくりへ貢献できると考えている。

外部リソースとのマッピングについては先行研究で機能語彙体系間でのマッピングにおける考察と同様、概念定義の違いや、視点の違い、あるいは粒度の違いから対応関係に非対象性が生じることがわかっている[10]。今後、類似する分類軸によるグループマッピングに基づいた相互互換の利用を設けるなど関連付けの成功率を高める仕組みが必要となる。

4 Linked Data 技術を用いた応用システム開発にむけて

前節の結果から、異常状態知識の統合には、本異常状態オントロジーを骨格として用いて各種リソースと関連付けることが有効なアプローチであることがわかった。そこで、知識統合の実現を目指して、異常状態オントロジーを拡張し、Link Data 技術を用いた応用システムを試作することとした。

4.1 異常状態オントロジーの Linked Data 化

異常状態オントロジーは「法造」を用いて記述されている。そこで、異常状態オントロジーの Linked Data 化にあたり、法造が提供する OWL エクスポート機能を用い、ロールホルダー間の is-a 関係の表現をサポートしつつ、ロール理論の詳細表現を省略することで記述を単純化したモデル³を利用した[11, 12]。これにより、OWL 形式で出力された異常状態オントロジーにおいては、異常状態の is-a 階層は rdfs:subClassOf, attribute-of 関係⁴で表されていた異常状態の対象物(O)、属性(A)および属性値(V)は、owl:Restriction を用いて表される。

この OWL 形式と異常状態オントロジーと3章で述べた外部リソースのマッピング情報は、2.1 節で議論したように概念レベルが必ずしも一致しないことを考慮して、リソースの種類毎に medo:hpo や medo:pato といったプロパティを導入して表現した。

³ 法造のエクスポートメニューで「OWL(B)」と表記されているもの。

⁴ 法造ではスロットとして表現される。

これらのマッピング情報は、今後、新たな情報が追加されることを考慮して、異常状態オントロジーとは独立した RDF ファイルとして管理している。

4.2 疾患コンパスと異常状態オントロジーの連携

これまで我々は疾患オントロジーを構築し、疾患を構成する異常状態の因果連鎖（疾患連鎖）に関する知識を Linked Data 化し、外部リソースと連携して疾患情報を閲覧できる Web アプリケーション「疾患コンパス」を公開している⁵[13]。そこで、本異常状態オントロジーの Linked Data と、疾患コンパスにおける疾患知識を関連付けることとした。

疾患オントロジーにおける疾患定義は、異常状態オントロジーで定義された異常状態概念を参照して定義されている。よって、異常状態オントロジーにおける異常状態を、疾患コンパスにおけるレベル 3 の疾患を構成する異常状態に対応づけた。技術的には、疾患コンパスが提供する API を用いて連携させることとした。

4.3 応用システムの試作と実装

3.1 節で述べた異常状態オントロジーの OWL ファイルおよび外部リソースとのマッピング情報を、トリプルストア⁶に格納し Linked Data として公開した。本 Linked Data の閲覧には、SPARQL クエリになじみがない生命医学研究者が容易に利用できるように、専用の閲覧システムを開発した⁷（図 3）。

本システムは、Java Script を用いた Web アプリケーションとして実装されている。システム上に表示される異常状態知識は、すべて、SPARQL クエリにより動的に取得されている。合わせて、疾患コンパスで公開されている疾患知識、マッピング情報を介した外部リソースと連携し、統合された異常状態知識を利用した応用システム開発に向けた、基本機能を実装した試作システムと位置づけられる。

4.4 試作システムの機能

試作システムでは、入力された用語について、異常状態オントロジー内部データセットの検索、疾患コンパスへのリンク情報、および外部リソースへのリンクが表示される。図 3-①に示すように、「異常状態 (Abnormal state)」、「対象物 (Target Object)」、「属性 (Attribute)」、「属性値 (Value of Attribute)」を条件として指定して検索ができる。検索対象は、さらに構造関連異常 (Structural Abnormality)、機能関連異常 (Functional Abnormality) などから異常状態をフィル

タリングすることができる（図 3-②）。また、異常状態の is-a 階層の表示（図 3-③）から異常状態を選択することも可能である。検索結果（図 3-④）については、いずれかを選択すると図 3-⑤のように、異常状態の定義内容表示ペインで、異常状態の定義内容および外部リソースおよび疾患コンパスの疾患情報を閲覧することができる。

以下では、これらの機能が異常状態知識の統合的利用に向けて、どのように貢献するかを議論する。

4.4.1 異常状態の共通性の把握

本システムでは、異常状態 is-a 階層の閲覧の他、異常状態および、異常状態の構成要素である対象物 (O)、属性 (A)、および属性値 (V) を用いることにより、観点に応じた検索を行うことができる。

例えば、ある器官で起こった異常状態を検索したい場合、対象物 (O) に探したい器官（例：冠動脈）を入力すると、「冠動脈閉塞」、「冠動脈壁肥厚」、「冠動脈狭窄」、「冠動脈攣収縮」のように器官固有の異常状態が結果として表示される。

次に、特定の属性に注目し、例えば「長さ(A)」に関する異常状態を集めよというクエリについては、「管内径縮小」などの他、「冠動脈壁肥厚」、「心室壁菲薄化」、「胃粘液層薄化」など長さのロール概念「長さ」も含めた結果を診療科横断的に取得することが可能である。

4.4.2 疾患コンパスとの連携による知識獲得

4.2 節で述べたように、疾患コンパスと異常状態オントロジーの連携を実現することで、複数疾患における異常状態の因果関係の共通性が把握できることがわかった。例えば、虚血の場合、疾患「脳梗塞」と「心筋梗塞」では、脳と心臓において部位は異なるものの、途中虚血までのパス（因果連鎖）はほぼ同じである（図 4）。いずれも、狭窄、閉塞などの構造的異常がおこるか、攣縮によって過収縮が起こり、（血流量が低下し）、虚血がおこり、虚血後の壊死が生じるという因果をたどることが共通する。その後、心筋壊死か、脳の細胞の壊死か部位によって下流が大きく異なることが疾患コンパスの因果関係によりわかる。前者は、心臓の心室の壁運動低下から、循環系に様々な症状を引き起こすが（図 4 上）、後者は脳から神経に影響を及ぼし、神経が支配する領域で麻痺など、循環器の場合とは異なる固有の状態を引き起こす（図 4 下）。

⁵ <http://lodc.med-ontology.jp/>

⁶ AllegroGrpah(<http://franz.com/agraph/>)を利用

⁷ <http://lodc.med-ontology.jp/> 閲覧システムの公開サイト

Search Abnormality which includes
虚血

in Abnormal state

Search

Abnormal States Filter

- Structural Abnormality
- Functional Abnormality
- Other Abnormality
- Parametric Abnormality
- Behavior / Motion
- Phenomenal Abnormality

☆一過性脳虚血【脳における血液供給相対不足】
☆一過性心筋虚血【心筋における血液供給相対不足】
脳虚血【脳における血液供給相対不足】
虚血【血液供給相対不足】
心筋虚血【心筋における血液供給相対不足】

- 属性値関連異常【パラメーター異常】
 - ⊕ 属性値関連異常【絶対・単純】
 - 属性値関連異常【相対】
 - 属性値関連異常【過不足】
 - 属性値関連異常【不足】
 - 時間当たり供給不足
 - 虚血【血液供給相対不足】
 - 脳虚血【脳における血液供給相対不足】
 - ☆一過性脳虚血【脳における血液供給相対不足】
 - 心筋虚血【心筋における血液供給相対不足】
 - ☆一過性心筋虚血【心筋における血液供給相対不足】

図4の上部は、循環器系症状のオントロジーを示しています。中心には「心筋虚血」があり、これに関連する概念として「狭窄」（大動脈、冠動脈）、「攣縮」（冠動脈、心筋）、「心筋梗塞」、および「心筋細胞壊死」が示されています。また、「循環器系症状」もこのネットワークの一部として位置づけられています。

Label	心筋虚血【心筋における血液供給相対不足】
Target Object(O)	心筋血液
Attribute(A)	定量的属性
Value of Attribute(V)	不足_v
ParentClass	虚血【血液供給相対不足】
PATO	
HFO	
MeSH	D054059
SNOMED-CT	414795007
Diseases included in Core Chains	
Diseases included in Derived Chains	虚血性心疾患 心筋梗塞

図4の下部は、神経症状のオントロジーを示しています。中心には「脳虚血」があり、これに関連する概念として「狭窄」（脳・頸部動脈の閉塞・狭窄）、「攣縮」（脳・頸部動脈の攣縮）、「脳梗塞」、および「脳細胞壊死」が示されています。また、「神経症状」もこのネットワークの一部として位置づけられています。

Label	脳虚血【脳における血液供給相対不足】
Target Object(O)	脳血液
Attribute(A)	定量的属性
Value of Attribute(V)	不足_v
ParentClass	虚血【血液供給相対不足】
PATO	
HFO	902637
MeSH	D002545
SNOMED-CT	287731008
Diseases included in Core Chains	
Diseases included in Derived Chains	虚血性脳血管障害 脳梗塞

図4 異常状態オントロジーと疾患コンパスの連携例

4.4.3 異常状態の外部リソースとの関連付け

今回対象とした12疾患で検証した外部リソース4つとマッピングを行ったものについては、OBO Foundry ライブラリ内各種オントロジーの Linked Data サーバーブラウザである Ontobee⁸ または外部リソースが提供するサイトの対応データにアクセスすることが可能である。例えば、試作システムの定義内容表示ページの MeSH ID をクリックすると、米国医学図書館のサイトの該当用語の情報にアクセスすることが可能である。

4.5 統合基盤構築の実現にむけて

本オントロジーとの統合によって、外部リソースのユーザーにとっては次の利点があると考えられる(図5)。

4.5.1 基礎から臨床へ異常状態知識の統合

本研究では、異常状態オントロジーを介して、生物種をヒトに特化した表現型 HPO と生物種横断表現型 PATO との統合を試みたが、今後、他の実験動物との関連付けを進めたいと考えている。生物種を超えたマウスやラットとの表現型とヒトの表現型を対応付けについては、属性(A)とその値(V)のセットが一致した上で、対象物(O)である解剖構造物同士の類似性を確かめなければならない。現在、マウス、ラットの表現型データベースの統合が試みられているので[14]、それらを利用し、Linked Data として関連付けることで生物種横断的なマッピングが可能になり、ヒトの疾患で起こる各種症状から実験動物で観察される概念までを統合化した知識を提供できると期待される。

また、本研究で統合を試みた MeSH は、米国国立医学図書館が提供する生命医学文献情報のアノテー

⁸ <http://www.ontobee.org>

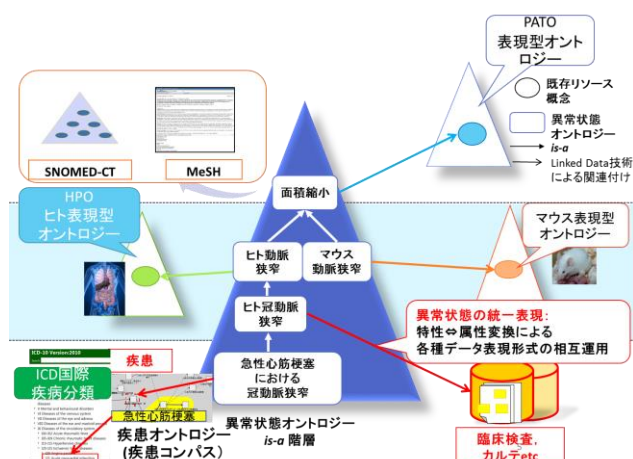


図5 異常状態オントロジー-Linked Data化と知識統合基盤の全体像

ションに用いられているため、関連文献情報の取得が可能となる。例えば、基礎で得られた異常状態に関する新知見を文献から得て、臨床で希少疾患の病態解析などに役立てるなどの応用が考えられる。

4.5.2 臨床的な意義

次に、臨床上的有用性について述べる。疾患については、WHOが提供する国際疾病分類ICD(International Classification of Diseases)⁹との関連付けを予定している。本分類は、国際死因分類に関する統計の他、近年は各種統計調査や医療機関における診療録の管理等に使われている。現在が第10版(ICD 10)であり、2017年改定予定の第11版ICD11では、疾患にその「原因因子(Causal Properties)」を記述する予定となっているので、将来、それらの中の異常状態とマッピングできればと考えている。ICD10/11の各疾患データと疾患オントロジーの疾患概念を対応づけることで、ICDユーザーにとっても、関連付けを介して注目した疾患でどのような異常状態の因果関係が起きているかを把握できる。

さらに、異常状態の対象となる対象物(O)の器官について、本研究グループの開発した解剖学オントロジーとの関連付けを予定している。その結果、各疾患で起こりうる異常状態について、隣接臓器や血管の接続情報などを知ることができ、影響する器官の状態予測や解析にも有用であると考えられる。

また、検査データについては、LOINC(Logical Observation Identifiers Name and Codes)¹⁰や国内では医療情報開発センターの提供するMEDIS¹¹標準マス

ターが標準用語やコードの提供を行っており、将来これらと本異常状態オントロジーの対応する各構成要素をマッピングすることを予定している。このように、モデル動物のような基礎から、臨床データ、さらに高次の異常状態知識まで疾患に関わる様々な知識についてその関係性が把握できることから、疾患の全体像の理解に貢献すると期待される。

5 まとめ

本研究では、異常状態を中心とした疾患知識統合のための課題をオントロジー工学の観点から述べるとともに、その解決法として異常状態オントロジーの概念レベルに応じた外部リソースとのマッピングについて提案した。さらに、統合実現に向けて異常状態オントロジーの応用として、Linked Data化を行い、知識統合に向けたシステム開発の取り組みについてのべた。

Linked Dataは、統合が成功すれば、自身のデータ拡充にもつながるといえる。しかし、表層ラベルのみの一致で関連付けを行うと、意味内容は一致しないことも多い。また、データ源にアクセスしてもその概念定義や属性が明示されていない場合があり、概念階層構造にエラーが生じている可能性もある。そこで、データに健全性を与え、知識統合の骨格となるのがオントロジーである。オントロジーは知識が明示され、客観的かつ一貫性のある本質属性を捉えた厳格な階層構造により、質の高い概念を保有する。オントロジーのLinked Data化は、誰もが簡単にオントロジーにアクセスし、参照するために有効な手段となる。Linked Data化したオントロジーを骨格として参照すれば、汎用レベルから専門ドメインまでの知識を自由に横断して、各種リソースの関連データについて信頼性のある情報取得が可能となる。

今回、統合試行において3診療科を横断した知識を把握することができたが、今後スケールアップすれば、もっと多くの診療科を横断した複数の疾患の異常状態の共通性が明らかになると考えている。

生命科学分野は、他分野に先行してLinked Data化が活発に進められており、その中で、複雑で多様な臨床情報をどう扱い、基礎研究と統合するかは、今後他のドメインでも直面する問題にも共通するものが多い。本研究で疾患や異常状態知識の統合が実現すれば、工学分野での人工物の故障や異常状態知識にも活用できると考えている。

謝辞

本研究の一部は、日本学術振興会の最先端研究開

⁹ <http://www.who.int/classifications/icd/en/>

¹⁰ <http://loinc.org/>

¹¹ http://www.medis.or.jp/4_hyojyun/medis-master/

発支援プログラムおよび厚生労働省医療知識基盤研究開発事業により、助成を受けたものである。これらの研究・開発に参加された方々（システム実装を担当いただいた加藤敦丈氏，他）に感謝致します。疾患オントロジー構築および臨床医学的な知識の提供を頂いた東大病院の臨床医の先生方（大友夏子先生，林亜紀先生，松村貴由先生，桜井亮太先生，寺田さとみ先生，脇嘉代先生），東大病院企画情報運営部 河添悦昌先生，梶野正幸先生，疾患生命工学センター 篠原恵美子先生に心より感謝を申し上げます。

参考文献

- 文集（第22回），1-4，(2008)
- [1 1] K. Kozaki, E. Sunagawa, Y. Kitamura and R. Mizoguchi.: Role Representation Model Using OWL and SWRL, Proc. of 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies, Berlin, July 30-31, pp.39-46. (2007)
- [1 2] 周俊，国府裕子，古崎晃司，溝口理一郎：OWL を用いたロール概念表現モデルに関する考察，2008年度人工知能学会全国大会（第22回），3G2-1，(2008)
- [1 3] 古崎晃司，山縣友紀，国府裕子，今井健，大江和彦，溝口理一郎：医療知識基盤の構築に向けた疾患オントロジーの LOD 化，人工知能学会誌，Vol. 29, No. 4, pp.396-405, (2014)
- [1 4] 榎屋啓志，古崎晃司，溝口理一郎：コンテキストに依存した定性値を扱う生物表現型統合データベースの試作，2013年度人工知能学会全国大会（第27回）資料集，3I1-2，(2013)
- [1] PL Whetzel, NF Noy, NH Shah, PR Alexander, C Nyulas, T Tudorache, MA Musen., BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res*, 39(Web Server issue):W541-5, (2011)
- [2] 溝口理一郎 オントロジー工学，オーム社，(2005)
- [3] 大江和彦，今井健：臨床医学知識処理を目指した医療オントロジー開発，人工知能学会誌，Vol. 25, No. 4, pp.493-500, (2010)
- [4] R. Mizoguchi, K. Kozaki, H. Kou, Y. Yamagata, T. Imai, K. Waki and K. Ohe: River Flow Model of Diseases, In Proc. of 2nd International Conference on Biomedical Ontology (ICBO2011), pp.63-70 July 28-30., Buffalo, NY, USA, (2011)
- [5] 山縣友紀，古崎晃司，今井健，大江和彦，溝口理一郎：疾患における異常状態オントロジーの構築，医療情報学，Vol. 34, No. 3, pp.101-117, (2014)
- [6] Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D: Using ontologies to describe mouse phenotypes, *Genome Biol*, 6, R8. (2005)
- [7] Köhler S et al.: The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data, *Nucleic Acids Res*, 42 (Database issue):D966-74. (2014)
- [8] Grenon P, Smith B, Goldberg L. Biodynamic Ontology: Applying BFO in the Bio-medical Domain, *Stud Health Technol Inform*, Vol. 102, pp.20-38,(2004)
- [9] Yamagata Y, Kozaki K, Mizoguchi R, Imai T, Ohe K. An ontological modeling approach for abnormal states and its application in the medical domain. *J. Biomed. Sem*, Vol.5 : 23, (2014)
- [1 0] 瀬川翔，大久保公則，笹嶋宗彦，來村徳信，溝口理一郎：機能オントロジーマッピングのガイドラインに関する一考察：2つの機能語彙体系間でのマッピングを例として，2008年度人工知能学会全国大会論