

生命科学系 LOD と異常状態オントロジーの統合

— 疾患知識横断的把握および相互運用に向けて —

An Integration of Abnormality Ontology with Biomedical Knowledge from LOD

山縣 友紀^{*1}
Yuki Yamagata古崎 晃司^{*1}
Kouji Kozaki今井 健^{*2}
Takeshi Imai大江 和彦^{*2}
Kazuhiko Ohe溝口 理一郎^{*3}
Riichiro Mizoguchi^{*1} 大阪大学産業科学研究所
The Institute of Scientific and Industrial
Research (ISIR), Osaka University^{*2} 東京大学大学院医学系研究科
Department of Medical Informatics,
Graduate School of Medicine,
The University of Tokyo^{*3} 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and
Technology

In this paper, we discuss applying Linked Open Data (LOD) techniques to abnormality ontology. We have developed a three-layer ontological model of abnormal states. This article focuses on linking abnormal states to several biomedical ontologies and terminologies appropriately. Then, we discuss integrated knowledge acquisition and interoperability between generic and human disease-dependent abnormal states.

1. はじめに

近年、生命科学分野において、これまで蓄積されたデータをオープンにし、情報を共有するという動きが高まっており、ここ数年、公開されるデータセット数は爆発的に増加している[Whetzel 11]. 一方、医療分野においては、現在、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、蓄積された様々な医療情報をいかに交換、知識共有し、さらに相互運用できるかが課題となっており、データから、疾患や異常状態というような概念レベルの高い知識までを統合的に扱えるような基盤の構築が強く求められている。

筆者らは、これまで約 7 年間にわたり臨床医学オントロジーの構築を進めており、オントロジー工学者と臨床医を含む医療領域の専門家の綿密な協力の下、疾患オントロジー、解剖学オントロジー、および異常状態オントロジーの体系化を行ってきた[大江 10, Mizoguchi 09, 11]. これらの概念を既存のオントロジーや標準語彙などと関連付けて、統合化することは上述のような知識の共有と活用において重要となる。

そこで、本稿では、現在構築中である異常状態オントロジー[山縣 12, Yamagata 12]と外部で公開されているデータセット(外部リソース)を Linked Open Data (LOD) 技術を用いて統合する試みを紹介するとともに、生命科学分野の異常状態知識統合への貢献について議論する。

以下、2 章では、本研究で LOD 化の対象とする異常状態オントロジーの概要を述べる。3 章では、異常状態オントロジーの各概念レベルに応じた外部リソースとの関連付けについて考察し、続く 4 章では本研究を総括すると共に今後の展望を述べる。

2. 異常状態オントロジーの概要

2.1 異常状態の3階層

異常状態の is-a 階層を構築するにあたり、一貫した視点で概念化を行うことが必要である。そこで、本オントロジーは、異常状態概念を下記のような 3 層構造として捉えることとした(図 1)。

- レベル1: 汎用異常状態(対象物非依存の異常状態)

連絡先: 山縣友紀, 大阪大学産業科学研究所, 〒567-0047 大阪府茨木市美穂が丘 8-1, Tel: 06-6879-8416, Fax: 06-6879-2123, E-mail: yamagata@ei.sanken.osaka-u.ac.jp

- レベル2: 対象物に応じて共通に現れる異常状態
- レベル3: 特定のコンテキスト(疾患)を構成する異常状態

なお、2014 年 1 月 14 日時点で本疾患オントロジーには、13 診療科に渡る 6,302 の疾患と、その疾患を構成する 21,669 の異常状態が定義されており、本稿ではそのうち、現在洗練作業が行われている 5 診療科、1,602 疾患を構成する 9,985 の異常状態を対象とする。

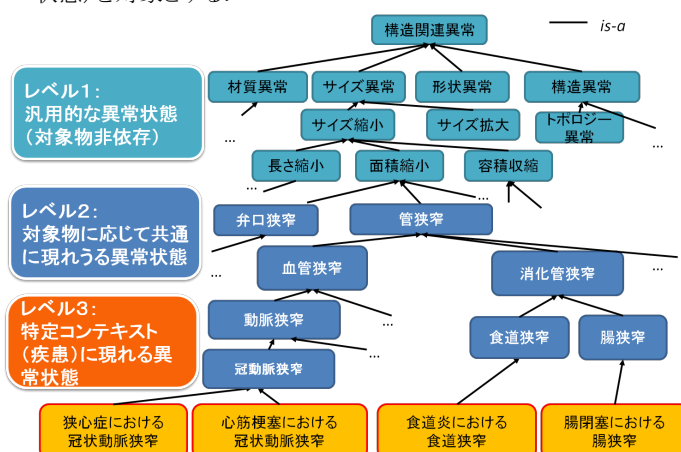


図1 異常状態オントロジー3階層モデル

2.2 異常状態オントロジーの計算機表現

異常状態オントロジーにおける疾患定義の「法造」による計算機表現の例を図 2 に示す。これは、異常状態を汎用レベルから疾患定義における異常状態までの特殊化の具体例を示したものである。

レベル1の汎用的異常状態である「面積縮小」は、属性「面積」とその値が「小さい」として定義されている(図 2 右上段)。

レベル2最上位(図 2 左上段)では汎用的異常状態「面積縮小」をクラス制約として参照し、対象物のクラス制約を「管状構造物」と特定するとともに上位から継承された「面積」を「断面積」に特殊化し、「管の断面積が小さい」と表現することで、管の異常状態である「管の狭窄」状態を定義することができる。また、レベル2の下位にいくにつれて、人体の解剖構造物である「血管」における「血管の狭窄」(図 2 左中段)、「動脈狭窄」(図 2 では省

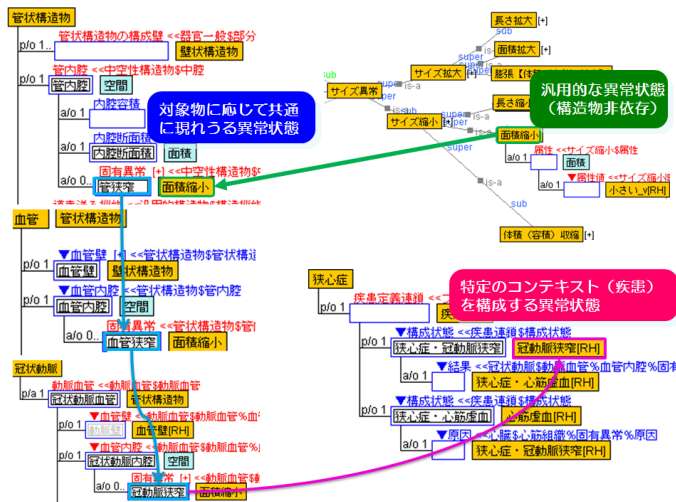


図2 異常状態オントロジーの計算機表現

略)、「冠動脈狭窄」(図2左下段)のように特殊化され、器官固有の異常状態として、必要とする粒度まで表現している。

さらに、「冠動脈狭窄」は、レベル3の疾患を構成する異常状態として、右下に示すように、例えば狭心症を構成する異常状態として「狭心症・冠動脈狭窄」のように特殊化される(図2右下段)。

本異常状態オントロジーは、汎用性をもたせたドメイン非依存かつ構造物非依存の異常状態から特殊化し、汎用的構造物依存として例えば管に共通する異常を定義し、さらに医療ドメイン内でも汎用的レベルから、特定コンテキストを構成する異常状態まで特殊化していくという、共通性と特殊性を明確に分離しながらコンテキストに応じた異常状態を表現することができる記述となっている[山縣 12, Yamagata 12]。

3. 生命医学分野各種リソースとの統合

3.1 異常状態と各種リソースの関連付け

異常状態概念は多種多様であり、外部リソースとの統合を考えた場合、これらの概念階層にあわせて適切に関連付けを行うことが必要となる。

ここで、生命科学分野では、実験による観測や臨床所見を通じて異常状態知識を把握するのが一般的であり、表現型を強く意識した形で異常状態に関する知識が存在する。そこで、まず、表現型のオントロジーとしてよく用いられている外部リソースとマッピングすることとし、具体的には、PATO (Phenotypic Quality Ontology) [Gkoutos 05], HP (The Human Phenotype Ontology) [Köhler 14] を対象として選定することとした(表1)。

表1 異常状態オントロジーと各種リソースの対応関係

臨床医学 オントロジー	概念レベル	概念	例	外部リ ソース	外部リソース公開サイト
異常状態オ ントロジー	レベル1	汎用異常状 態	面積縮小	PATO	http://obofoundry.org/wik/index.php/PATO:Main_Page
	レベル2	対象物依存 異常状態	動脈狭窄	HP	http://www.human-phenotype-ontology.org/
	レベル3	疾患依存異 常状態	冠動脈疾患に おける冠動脈 動脈狭窄	(ICD11)	http://www.who.int/classifications/icd/revision/en/
疾患オント ロジー		疾患	虚血性心疾患	ICD10	http://www.who.int/classifications/icd/en/

前章で述べたように、本異常状態オントロジーは3層構造で体系化している。最上位(レベル1)の異常状態は、複数の対象物に汎用的に見られる(対象物非依存)共通の異常状態(汎用異常状態)を定義しており、構造関連異常、機能関連異常、パラメーター異常などに大別されている。概念例として、圧力増加、変形、機能低下等が挙げられるが、これらはバイオメディカル分野以外でも用いられる一般的な基本概念といえる。

一方 PATO は、バイオ分野で表現型や性質を扱うオントロジーとして知られており、生物種を横断した概念を提供し、バイオコミュニティで広く使われている。例えば、「部品の欠損」、「面積縮小」、「圧力増加」のように対象生物種を限定せず汎用的な概念を体系化しており、これらは本異常状態オントロジーのレベル1汎用異常状態(対象物非依存の異常状態)と対応すると考えられる。そこで、本レベル1の異常状態概念を PATO の対応概念へマッピングすることとした。

次に、異常状態オントロジーのレベル2では、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の最上位では、複数ドメインに共通する形態の全体物を対象とした異常状態が定義される。形態の全体物とは、形態そのものにアイデンティティをもつオブジェクトであり、例えば、中空性構造物、袋状構造物などを用いる。レベル1で定義された汎用的な異常状態を異常対象に沿って特殊化することで、レベル2の異常状態の is-a 階層が定義される。

レベル2の更に下位に行くにつれて、対象となるドメインのオブジェクト(対象物)下で現れる共通の異常状態を担う概念が定義されていく。本異常状態オントロジーでは、人体解剖構造物を対象として体系化しており、レベル2最上位の管状構造物は、人体固有の「人体管状構造物」コンテキストで先の「管狭窄」を特殊化した「人体管狭窄」となって定義される。さらに、それが血管で現れると「血管狭窄」となり、冠動脈では「冠動脈狭窄」となるように、専門医が必要とする粒度まで階層記述となっている。

一方 HP はヒトの表現型に特化したオントロジーとして知られており、遺伝病 DB である OMIM[OMIM]の記述に際し、登場する表現型に関して統制語句を提供するために開発されたものである。現在、約 10,000 用語が公開されており、例えば「肺高血圧」、「冠動脈狭窄」のような各器官でおこる異常状態が定義されている。これらの概念は、上述した本異常状態オントロジーのレベル2の異常状態の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の異常状態概念を HP の対応する概念に関連付けることとした。

さらに、本異常状態オントロジーでは、レベル3に疾患コンテキスト依存の異常状態について定義している。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」などがあげられる。しかし、このような、疾患コンテキストに依存する異常状態概念は外部リソースではあまり存在しない。

しかし、WHO が提供する 2017 年改定予定の国際疾病分類 ICD11[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。

なお、本異常状態オントロジーではレベル3では疾患コンテキスト依存の異常状態を専門医が定義しており、例えば「排尿後の一過性の血圧低下」のように、その疾患で何が起こったかを専門医の高度かつ専門性のある特殊化した概念として詳細に記述されており、修飾語が多く用いられている。これら修飾語は、修飾属性として別に扱っており、それら修飾語を除いた概念を外部リソースと対応付ける予定である。

本オントロジーとこれら表現型リソースとの統合によって、外

部リソースのユーザーにとっても次の利点があると考えられる。

外部リソース利用者は、適当な概念について、マッピングされた本オントロジーツリーをみれば、その概念の汎用レベルから特殊化された階層までのレベルを把握できる。例えば、HP は、どこでその表現型が起こっているかという解剖構造からの視点で表現型概念を体系化している。そのため、動脈狭窄 *is-a* 血管異常というように「狭窄」という異常状態自体の体系化は不明である場合が多い。そこで、例えば、本異常状態オントロジーの「動脈狭窄」とマッピングしていた場合、その HP の「動脈狭窄」をみたユーザーは、「動脈狭窄 *is-a* 血管狭窄 *is-a* 管狭窄 *is-a* 面積縮小」という階層構造を本オントロジーによって把握することが可能となる。さらに、面積縮小は生物種を横断した表現型 PATO と結びついていることがわかる。つまり、本オントロジーを介して、生物種をヒトに特化した表現型 HP と生物種横断表現型 PATO とそれらとの関係が把握できることになる。

なお、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 similarity をいなければならない。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[梶屋 13]ので、それらを利用し、LOD としてつなぐことで生物種横断的なマッピングが可能になり、ひいては生物種横断概念から各種生物種に特化した概念までを統合化した知識を提供できると考えている。

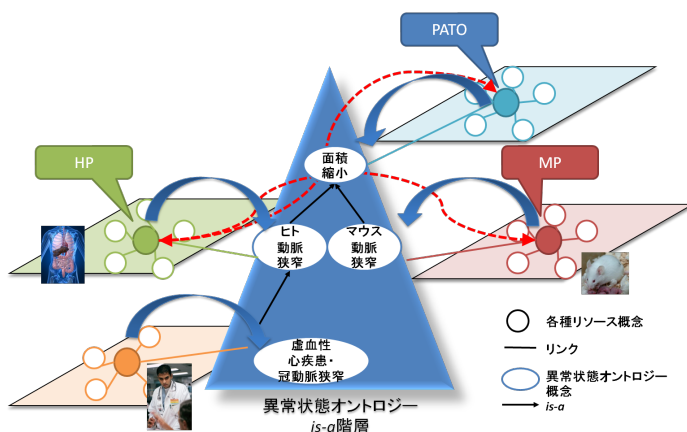


図3 異常状態オントロジーと表現型 LOD 統合

3.2 異常状態の各構成要素と各種リソースの関連付け

本異常状態オントロジーにおいて、異常状態は特性(P)で表現し、それを属性分解したトリプル表現<対象物(O), 属性(A), 属性値(V)>で表現することとしている。そこで、これら構成要素を各種外部リソースと対応付けることを検討している。例えば、FMA (The Foundational Model of Anatomy ontology) [FMA] は解剖学オントロジーとして著名なオントロジーであり、レベル2の異常状態の対象物を FMA とマッピングする予定である。また、検査データについては、LOINC[LOINC]や国内ではMEDIS[MEDIS]が標準用語、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で、データと疾患定義における異常状態のような高次知識との相互運用に貢献すると考えている。

3.3 疾患と ICD10 の関連付け

ICD10 は、WHO が提供する国際疾病分類第 10 版であり、もともとは国際死因分類に関する統計のために設けられたが、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[ICD10]。

そこで、本疾患オントロジーの各疾患概念と ICD10 に記述されている疾患名に対応付けることとした¹。

ここで、疾患オントロジーの疾患連鎖では、各疾患を構成する異常状態(レベル3)の因果関係を専門医によって定義されているので、ICD10 のユーザーにとっても、本オントロジーとのリンクを介して、その疾患がどのような異常状態の因果関係が起こっているかを理解するのに有用であると考えている。

さらに、例えば ICD10 の狭心症で起こりうる異常状態冠動脈狭窄について、HP とひもづいているので、それら外部リソースを介してヒトの表現型および OMIM との対応関係も可能になる。また、前節で述べたように異常状態について、対象物を規定する解剖学オントロジーと FMA のリンクがあれば、ICD10 コードに対応する疾患が起こりうる解剖構造に関する情報を FMA コードも含めて知ることができるという利点がある。

3.4 LSD, MeSH との関連付け

検索をする上で、本異常状態オントロジーで記述されている概念と一般的に用いられている辞書の用語とを対応付けを行うておくことは有用であると考えられる。そこで、生命科学分野で研究者に広く用いられているライフサイエンス辞書(LSD)[LSD]とのマッピングを行う予定である。ライフサイエンス辞書の各用語は、米国国立医学図書館(NLM)が提供する生命科学分野の文献の統制語彙である MeSH ターム[MeSH]との対応付けが行われている。MeSH タームには、疾患および異常状態に関する概念も含まれるため、LSD の概念を介して MeSH タームとリンクしておけば、異常状態および疾患に関する文献の情報との関連付けが可能となると考えている。その結果、例えば、疾患連鎖のある異常状態間の因果関係に関する文献を集め、因果関係について解析を行い、既に記述した因果関係について更なる知識の追加や、別の因果関係について新たな知見の取得などへ応用できる可能性がある。また、文献を記述した研究者にとっても、オントロジーを介してその異常状態が別の疾患でどのような異常状態と関係するか、また、オントロジーの兄弟概念として、他の臓器で起こっている同様の異常状態を検出したり、生物種を超えて関連する異常状態知識を把握するなどさらなる研究の発展に役立つ可能性がある。

4. まとめおよび今後の検討課題

これまで、本研究グループでは、疾患オントロジーから疾患連鎖に関する情報のみを抽出してRDF化した疾患連鎖LOD²を公開している[古崎 13]。今後、異常状態オントロジーをLODとしてどのように公開するか現在検討を行う予定である。

非常に高度かつ専門性のある特殊化した概念は、ともすればそのドメイン内の非常に狭い範囲のみの知識として完結してしまうおそれがある。一方、非常に一般的な概念は実用上使いにくいという側面もある。そこで、特殊化された概念から汎用的な概念まで、概念層に切り分けて、その概念レベルに適切に応じた外

¹ 本研究では ICD10 分類の第 1 章-14 章を対象とした

² <http://lodc.med-ontology.jp/>, 疾患連鎖 LOD の公開サイト

部リソースとマッピングすることで、臨床医学ドメイン知識から生物種を超えた知識まで、生命科学分野の異常状態に関連する知識を統合することができると考えている。

このように、LOD とオントロジー組み合わせることにより、すでに構築された概念を積極的に利用し、自身の専門以外の分野の知識について理解を深めることができるとともに、分野を超えたユーザーの多様な視点に基づいた各研究にも利用できるような相互運用性の向上に有用であると考えている。その結果、基礎研究と臨床研究の橋渡しとなり、生命科学分野全体の知識基盤構築に貢献することが期待される。

謝辞

本研究の一部は、日本学術振興会の最先端研究開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。疾患オントロジーの記述と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、脇嘉代先生ら)、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の感謝を申し上げます。

参考文献

- [Whetzel PL 11] PL Whetzel, NF Noy, NH Shah, PR Alexander, C Nyulas, T Tudorache, MA Musen., *Nucleic Acids Res*, 39(Web Server issue):W541-5, (2011).
- [大江 10] 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.
- [Mizoguchi 09] R. Mizoguchi, H. Kou, J. Zhou, K. Kozaki, T., Imai and K. Ohe: An Advanced Clinical Ontology, In *Proc. of International Conference on Biomedical Ontology (ICBO)*, pp.119-122, Buffalo, NY, June 24 - 26, (2009)
- [Mizoguchi 11] 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, 2011, Buffalo, NY, USA, 2011.
- [山縣 12] 山縣友紀, 国府裕子, 古崎晃司, 今井健, 大江和彦, 溝口理一郎:異常状態オントロジーとその応用, 2012 年度人工知能学会全国大会(第 26 回)資料集, 112-R-4-3, 2012.
- [Yamagata 12] Y. Yamagata, H. Kou, K. Kozaki, R. Mizoguchi, T. Imai, K. Ohe: Ontological Modeling of Interoperable Abnormal States, In *Proc. of the 2nd Joint International Semantic Technology Conference (JIST2012)*, LNCS 7774, pp. 33-48, Nara, Japan, Dec. 2-4, 2012.
- [Gkoutos 05] Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D: Using ontologies to describe mouse phenotypes, *Genome Biol*, 6, R8. (2005)
- [Köhler 14] 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)
- [OMIM] <http://www.ncbi.nlm.nih.gov/omim>
- [ICD11] <http://www.who.int/classifications/icd/revision/en/>
- [梶屋 13] 梶屋啓志, 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, 311-2, 2013.
- [FMA] <http://sig.biostr.washington.edu/projects/fm/>
- [LOINC] <https://loinc.org/>
- [MEDIS] http://www.medis.or.jp/4_hyojyun/medis-master/
- [ICD10] <http://www.who.int/classifications/icd/en/>
- [LSD] <http://lsd.pharm.kyoto-u.ac.jp/ja/index.html>
- [MeSH] <http://www.nlm.nih.gov/mesh/meshhome.html>
- [古崎 13] 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臨床医学オントロジーの LOD 化の検討, 2013 年度人工知能学会全国大会(第 27 回)資料集, 1N5-OS-10c-1, 2013.