

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

A Consideration of Building LOD from a Medical Ontology toward a Knowledge Infrastructure

古崎 晃司^{*1} 山縣 友紀^{*1} 国府 裕子^{*1} 今井 健^{*2} 大江 和彦^{*2} 溝口 理一郎^{*3}
 Kouji Kozaki Yuki Yamagata Hiroko Kou Takeshi Imai Kazuhiko Ohe Riichiro Mizoguchi

^{*1} 大阪大学産業科学研究所 ^{*2} 東京大学大学院医学系研究科 ^{*3} 北陸先端科学技術大学院大学
 The Institute of Scientific and Industrial Research (ISIR), Osaka University Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo Japan Advanced Institute of Science and Technology

This article discusses a construction of Linked Open Data (LOD) from a medical ontology to use it as a knowledge infrastructure. Because the medical ontology has been developed under a close cooperation between ontology engineer and medical experts, it could be a very valuable knowledge base for advanced medical information systems. This article focuses on a disease ontology in which diseases are defined as causal chains. We consider how design a RDF data model for diseases as causal chains from a viewpoint of convenience as LOD. Then, we show a prototype of disease chain LOD as the result.

1. はじめに

近年、医療分野における情報化は世界的に進められており、電子カルテをはじめとする病院システムの情報化やそれに伴い蓄積された多量の情報の有効活用などが盛んに行われている。そのような医療情報システムの高度化を支える知識基盤として、医療分野を対象としたオントロジー構築が求められている[大江 10]。このような背景の下、筆者らは厚生労働省の支援を受け、オントロジー工学の成果を駆使した適切な概念化に基づく洗練された臨床医学オントロジーを目指し、オントロジー構築を進めてきた [古崎 08, Mizoguchi 09, 11]。その構築過程は、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で進められており、その成果となる臨床医学オントロジーは、深い専門性とオントロジー的な整合性を両立したものであると言える。

構築された臨床医学オントロジーには、13 診療科の代表的な約 6,000 の疾患定義を対象とした疾患オントロジー、人体の全血管、神経、筋肉および骨の接続関係の記述を含む、主要な解剖学構造を定義した解剖学オントロジーなどが含まれる。これらの成果を有効に活用し、様々な医療情報システムの開発を促進するには、臨床医学オントロジーをオープンな知識基盤として公開することが望まれる。そこで本研究では、近年、様々な領域におけるデータの公開手法として注目されている Linked Data 技術を利用し、臨床医学オントロジーを Linked Open Data (LOD)として公開することを検討する。

Linked Data は、形式上は RDF で書かれたデータであるため、OWL や RDF(S)などの RDF 形式で表現したオントロジーを Linked Data として公開することは容易にできる。しかし、OWL などのオントロジー記述言語は適切な推論処理を行うことを前提としたクラスレベルの記述を中心に設計されているのに対し、Linked Data ではインスタンス間のリンクを中心とした処理を中心に考えられている。そのため、OWL 形式のオントロジーをそのまま Linked Data として公開しただけでは、必ずしも、利便性に富んだ LOD になるとは限らない。そこで本研究では、疾患オントロジーを例として、臨床医学オントロジーの有効活用を促進す

るにあたって適切な LOD としての公開方法を検討する。

以下、2 章では、本研究で LOD 化の対象とする疾患オントロジーの概要を述べる。3 章では、疾患オントロジーの LOD 化の方針について考察し、疾患を構成する因果連鎖を対象とした LOD 化を行う。続く 4 章では臨床医学オントロジーを LOD 化するにあたっての今後の課題を考察し、5 章で本研究を総括すると共に今後の展望を述べる。

2. 疾患オントロジーの概要

2.1 疾患の概念定義

疾患を理解する上では、その疾患が何を起因としており、どのような結果(病態や症状)を引き起こすかを適切に捉えることが重要とされる。そこで本研究では、疾患概念を「その原因と途中経過を含めた一連の状態変化の連鎖と、それにより引き起こされている結果状態との総体」として捉える[大江 10, Mizoguchi 11]。すなわち、疾患は人体に現れる状態変化による因果連鎖から構成される従属的実在物として定義される。ここで因果連鎖は川の流れるように自らを変化させるものとして捉えることができ、オントロジー的には持続物の下位概念として定義できる[Mizoguchi 11, 古崎 11]。疾患定義に用いられる因果連鎖(以下、疾患連鎖と呼ぶ)は人体に現れる(異常)状態とその間の因果関係から成り、その種類は、

- **汎用疾患連鎖**: 人体で起こりうると思われる全連鎖。すべての疾患から共通して参照される。
- **疾患定義連鎖**: 疾患定義に必要な連鎖。その疾患を発症した患者には共通して見られる。
- **派生連鎖**: ある疾患定義連鎖から上流/下流方向に汎用疾患連鎖を辿って得られる連鎖。上流側はその疾患の原因と考えられうる連鎖、下流側はその疾患を発症した患者に起こりうる連鎖を意味する。

に分かれる。

疾患定義の中心は「疾患定義連鎖」であり、疾患の is-a 関係は、疾患定義連鎖の範囲の包含関係で表される¹。例えば、糖尿病の疾患定義連鎖が「インスリン作用不足→高血糖」であると

連絡先: 古崎 晃司, 大阪大学産業科学研究所, 知識システム研究分野, 〒567-0047 大阪府茨木市美穂が丘 8-1, Tel: 06-6879-8416, E-mail: kozaki@ei.sanken.osaka-u.ac.jp

¹ 疾患定義連鎖の範囲は変わらないが、それを構成する異常状態がより詳細な状態に特殊化されることにより下位疾患が定義されることもある。

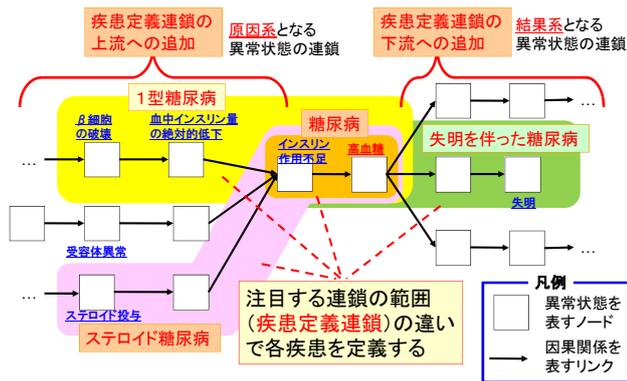


図1 糖尿病における疾患連鎖の例

すると¹, 「β細胞の破壊→血中インスリン量の絶対的低下→インスリン作用不足→高血糖」を疾患定義連鎖とする1型糖尿病は糖尿病の下位概念となる(図1)。すなわち, 疾患の下位概念の定義は, 疾患定義連鎖の範囲を派生連鎖に沿って上流・下流に広げたものとなる。

一方, 派生連鎖は, 定義には含まれないがその疾患で典型的に見られると思われる異常状態の連鎖を表す「疾患派生連鎖」と, すべての汎用疾患連鎖を辿ることのできる「一般派生連鎖」に分かれる。本研究では, 臨床医が自らの持つ専門知識と臨床の教科書の記述をもとに, 各疾患の疾患定義連鎖および一般派生連鎖を定義することで, 疾患オントロジーを構築した。

このように臨床医の手によって定義された各疾患を構成する因果連鎖(疾患定義連鎖と疾患派生連鎖)を一般化することで, 通常は, 定義することが現実的には不可能と思われる人体で起こりうる全ての因果連鎖(汎用疾患連鎖)を, ボトムアップ的に定義することができる。当然ながら, この方法で定義した汎用疾患連鎖は, 現実世界で人体に起こりうる全ての因果連鎖を網羅したものではないが, 本疾患オントロジーで定義した全疾患において臨床医が関心を持つ範囲で起こりうると思われる全連鎖に相当する。よって, この汎用疾患連鎖を用いることで, 疾患定義の際に直接記述した因果連鎖のみならず, 全診療科を横断した汎用疾患連鎖を用いた推論が可能となる。

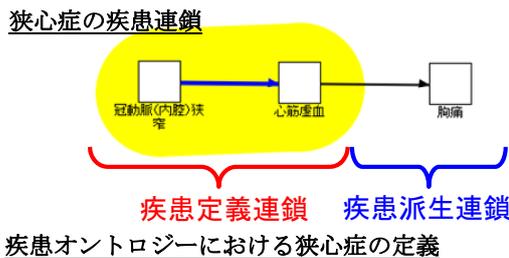
なお, 2013年3月11日時点で本疾患オントロジーには, 13診療科に渡る6,302の疾患と, その疾患を構成する21,669の異常状態が定義されている²。

2.2 疾患オントロジーの計算機表現

図2に疾患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示す。疾患(「狭心症」)の右に示された「因果構造物」と書かれた長方形は, 因果連鎖の一般構造を定義した概念への参照を示しており, 疾患が因果連鎖から構成されることを表している。疾患連鎖に現れる異常状態は「構成要素」として列挙され, それらが疾患定義連鎖/疾患派生連鎖のいずれに属するかは「連鎖種別」属性で表される。異常状態間の因果関係は「原因/結果」スロットの相互参照で表現されており, 必要に応じて発生確率属性が付与される。下位疾患の定義では, 上位疾患から継承された構成要素の特殊化や, 新たに疾患定義連鎖/疾患派生連鎖の範囲に入った異常状態の構成要素との追加が行われる。

¹ これは説明用に簡略化したものであり, 本研究の疾患オントロジーにおける糖尿病の定義とは異なる。

² このうち2,103疾患(13,910の異常状態)の定義が2010-2012年度のプロジェクトで大幅に見直されたものである。



疾患オントロジーにおける狭心症の定義

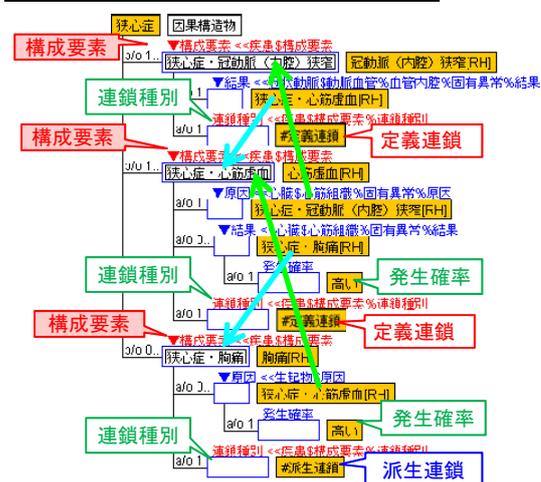


図2 疾患定義の計算機表現例

なお, 各構成要素は汎用連鎖として定義された異常状態を参照し, 各異常状態の定義からその疾患定義に必要な定義のみを継承することで, 疾患の種類に依らない汎用疾患連鎖との違いを明示している。すなわち, 疾患定義における構成要素は, 異常状態オントロジーの3階層モデルにおけるレベル3「特定のコンテキスト(疾患)を構成する異常状態」に相当し, レベル2「対象物内で汎用的に現れうる異常状態」を参照することで定義されている[山縣12, Yamagata12]。

3. 疾患オントロジーの Linked Open Data 化

3.1 LOD 化の基本方針

疾患オントロジーを LOD 化するにあたり, 最も容易な方法としては, 「法造」のオントロジーエクスポート機能を用いて, 疾患オントロジーを法造形式から OWL 形式のオントロジーに変換すればよい。しかし, OWL 形式に変換した疾患オントロジーをそのまま LOD として公開したとすると, 疾患の因果連鎖を取得するために複雑な SPARQL クエリを用いる必要がある。例えば, 異常状態 A から結果として起こりうる異常状態を辿る際には, 図3のような空白ノードを含む RDF グラフの検索を繰り返す必要がある。さらに, 図1のようなある疾患定義に含まれる因果連鎖を取得しようとすると, 図3のような構造を持つグラフの検索に加え, 上位疾患から継承された owl:Restriction を考慮した, 更に複雑なクエリが必要となる。これは, 結果として得られる図1のような因果連鎖を表すグラフ構造とは直観的に大きく異なり, LOD として公開した際の利便性を考慮すると, あまり好ましくな

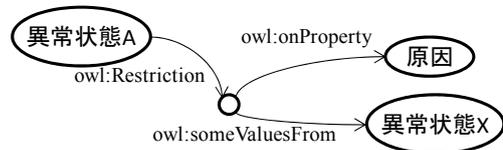


図3 OWLを用いた汎用疾患連鎖の表現

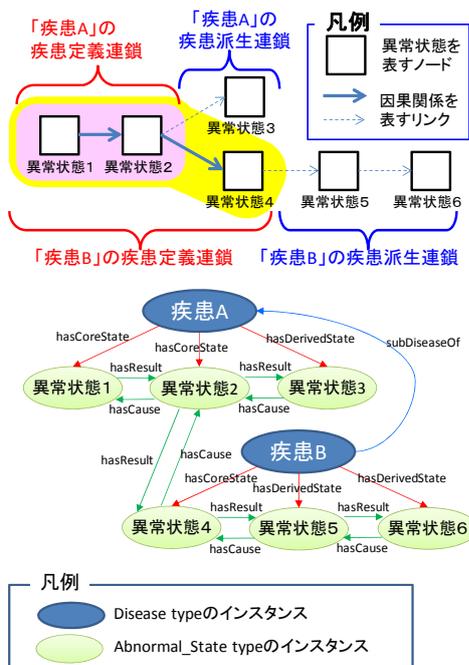


図4 疾患連鎖(上図)のRDF表現(下図)例

と思われる。

推論機構を用いて「ある因果連鎖のインスタンスが疾患オントロジー上で定義されたどの疾患に分類されるかを判定する」といった推論処理を行う際には、OWL形式のオントロジーが必要となる。しかし、「疾患オントロジーを疾患定義が体系化された知識源として参照するのみで、特定の疾患のインスタンスを対象とした処理をしない」ことを前提とすると、より簡易なRDF表現を用いることのメリットが大きいと考えられる。

そこで本研究では、後者の立場から、LODとしての利用の際の利便性を考えた、疾患オントロジーのLOD化を検討した。なお、初期的な検討を目的としたため、LOD化の対象は疾患オントロジーのなかでも中心となる疾患連鎖に限定した。

3.2 疾患連鎖 LOD の RDF データモデル

前節で述べた方針のもと、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LODを構築した。疾患連鎖LODにおいて、本来は「クラス(概念)レベル」の情報である疾患および異常状態の定義を、LODとして利用の際の利便性を考え、全てインスタンス(RDFリソース)に変換している。

異常状態は Abnormal_State タイプ(クラス)のインスタンスとして表現され、異常状態間の因果関係は hasCause および hasResult プロパティを双方向に記述することで表現される。これらのプロパティは、各異常状態の原因/結果となりうる異常状態を表し、それらを全ての異常状態から辿ることで汎用疾患連鎖が得られる。

疾患は Disease タイプのインスタンスとして表現され、疾患定義連鎖および疾患派生連鎖を構成する異常状態は、hasCoreState および hasDerivedState プロパティを用いて表される。また、通常のOWLやRDF(S)では rdfs:subClassOf を用いて is-a 関係が表されるが、疾患連鎖LODでは疾患がインスタンスとして表現されているため、疾患の is-a 関係を表すために、subDiseaseOf プロパティを導入した。

図4に疾患Aとその下位疾患Bの疾患連鎖(上図)をRDFで表現した例(下図)を示す。各疾患を構成する疾患定義連鎖および疾患派生連鎖を取得するには、各疾患との間に

図5 疾患連鎖 LOD の SPAQL エンドポイント。

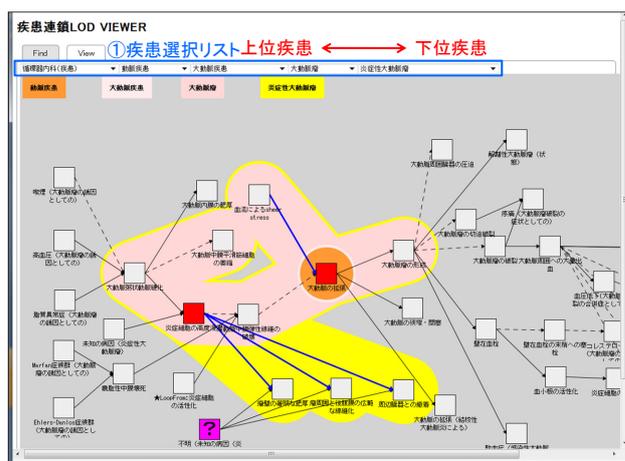


図6 疾患連鎖 LOD の可視化ツール。

hasCoreState および hasDerivedState プロパティを持つ異常状態を全て取得し、それらの異常状態間の hasCause/hasResult プロパティを辿ればよい。ここで、各異常状態は、2.2節で述べた疾患オントロジーにおける扱いとは異なり、「汎用疾患連鎖を表す異常状態」と「疾患の構成要素となる異常状態」は区別せずに、同一のインスタンスとして表されている。しかし、hasCause/hasResult プロパティでつながった異常状態が同じ疾患との間に hasCoreState/hasDerivedState プロパティを持つかどうかで、その因果関係がその疾患の疾患定義連鎖/疾患派生連鎖に含まれるかどうかの判定ができる。図4の疾患Aの場合、異常状態2の結果として異常状態4が記述されているが、異常状態4は疾患Aとの間に hasCoreState も hasDerivedState も持たないので、疾患Aの定義連鎖にも疾患派生連鎖にも含まれないことが分かる。

なお、疾患Bのように上位疾患を持つ疾患の疾患連鎖を取得する場合には、subDiseaseOfを辿って得られる上位疾患に含まれる疾患連鎖を合わせて取得する必要がある。このような処理を避け、上位疾患の有無に関わらず同じ方法で疾患連鎖を取得できるようにするには、あらかじめ全上位疾患から継承された構成要素となる異常状態に対して hasCoreState/hasDerivedState プロパティを持たせておくという方法が考えられる。しかし、この方法では、下位の疾患になればなるほど上位と同様に記述を多く持たせる必要があり冗長となるため、本疾患連鎖LODではこの方法を採用しなかった。

上述のように本データモデルでは、疾患連鎖の情報が、OWL形式に比べて直観的に表現されており、必要な情報を簡易なクエリで取得することができる。なお本データモデルでは、

現状の疾患オントロジーで定義されている疾患連鎖の情報は、全て表現されている¹。

3.3 疾患連鎖 LOD の試作と公開

前節で述べたデータモデルに基づき、疾患オントロジーから抽出した疾患連鎖を RDF 形式に変換し、LOD として公開した²。変換元の疾患オントロジーは代表的な 6 診療科を対象とした 2013 年 1 月 29 日時点のもので、約 1,800 の疾患定義とそれを構成する約 12,000 の異常状態の定義を含む。

合わせて疾患連鎖 LOD を外部システムから検索・利用するための API (SPARQL エンドポイント) と、疾患連鎖を図 1 のようなグラフ構造で可視化する疾患連鎖 LOD Viewer を開発した。SPARQL エンドポイント (図 5) には、初心者でもリストを選択するだけで簡単なクエリを試すことができる「簡易クエリ入力支援機能」と、検索結果から選択した RDF リソースを可視化する機能、および、検索結果に疾患が含まれる場合は「疾患連鎖 LOD Viewer」と連携して疾患連鎖を可視化する機能も実装した。

一方、疾患連鎖 LOD Viewer (図 6) は、

- 「疾患名」、「その疾患に含まれる異常状態名」から疾患定義を検索
- リスト形式で表示した「疾患の is-a 階層」(図 6-①) から表示する疾患を選択

の何れかの方法で選択した疾患の疾患連鎖を可視化することができる。本 Viewer は PC のみならず、タブレット端末、スマートフォンなど様々なデバイスから利用可能な Web サービスとして開発されている³。技術的には、疾患連鎖 LOD の SPARQL エンドポイントに対して選択した疾患の疾患連鎖の情報を取得するのに必要な SPARQL クエリを発行し、その結果を用いて可視化している。すなわち、本 Viewer で疾患連鎖が正しく表示されるということは、3.2 節で議論した方法に沿って正しく疾患連鎖の情報が取得できていることを示している。表示する疾患の種類やクライアントの環境にも依存するが、無線 LAN 接続下において疾患連鎖が表示されるまでに要する時間は、概ね 2 秒前後であり、現時点においては十分な速度であると思われる。

なお上述のシステムの開発にあたり、RDF データベースには AllegroGraph4.10 を可視化部分には HTML5 を利用している。

4. 臨床医学オントロジーの LOD 化に向けた考察

疾患オントロジーの LOD 化については、前章で述べたようにオントロジーにおける概念 (クラス) を RDF におけるインスタンスとして表現することによって、検索などの利用時に直観的かつ簡易なクエリで必要な情報を取得できるような LOD 化が可能であることを示せた。解剖学オントロジーにおける血管や神経の接続構造などについても、同様のアプローチが適用可能と思われる。しかしその際に、対象に応じて異なる RDF データモデルを検討することは、臨床医学オントロジー全体の LOD 化を考えた際に一貫性や整合性を乱す恐れがある。そのような問題を避けるために、対象とする概念の種類に依らず適用可能な、オントロジーにおける概念定義を RDF インスタンスに変換する汎用ルールを導入するという方法が考えられる。

その一方で、OWL や RDF(S) など既存のオントロジー記述言語をそのまま RDF としてもものを LOD として公開し、必要な情報

を取得するためのクエリの複雑さは、検索に用いる RDF グラフをパターン化することで吸収するというアプローチも考えられる。

いずれのアプローチを適用すべきは、公開した LOD の用途や規模等によっても変わるであろうが、臨床医学オントロジーのように大規模かつ多様な概念定義を含むオントロジーの LOD 化は、これらのアプローチの特徴を比較・検討するという点においても、重要な課題になると思われる。

特に、インスタンスを中心とした LOD に、リッチな概念定義を持つオントロジーを組み合わせた際に、知的な意味処理や知識の体系化といった観点で、どのような効果を得ることができるかを明らかにすることは、オントロジーを知的基盤として活用するための最重要課題の1つとして考察を深めたい。

5. おわりに

本論文では、疾患オントロジーを例として、臨床医学オントロジーを知識基盤として活用するために LOD として公開する方法について考察すると共に、疾患を構成する因果連鎖の LOD 化事例について述べた。公開された疾患連鎖 LOD は、Linked Open Data チャレンジ Japan2012 において「ライフサイエンス賞」を受賞し、臨床医学に基づいた知識を体系化した非常に有用なデータセットとして高い評価を得ている。今後、疾患以外のオントロジーを含めた臨床医学オントロジーの LOD 化についても、継続して検討を進めていく。

なお、臨床医学オントロジーの詳細については、Web サイト (<http://www.med-ontology.jp/>) に公開されている。

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。また、疾患オントロジーは、医療知識基盤研究開発事業に参加された方々の多大なる協力のものとて開発されたものであることに感謝致します。

参考文献

- [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.
- [古崎 08] 古崎晃司, 国府裕子, 周俊, 今井健, 大江和彦, 溝口理一郎: 臨床医学オントロジーの構築とその基本思想, 人工知能学会研究会資料, SIG-SWO- A802-09, 2008.
- [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 and et. al.: 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.
- [古崎 11] 古崎晃司, 他: 臨床医学オントロジーにおける疾患連鎖モデルの考察, 人工知能学会研究会資料, SIG-SWO-A1102-08, 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, I12-R-4-3, 2012.
- [Yamagata 12] Y. Yamagata and et.al.: 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.

¹ 紙面の都合で説明を省略しているが、構成要素となる異常状態の特殊化による、疾患定義連鎖の範囲が変わらない疾患の is-a 関係などより詳細な情報についても、isSpecializedFrom などといった別のプロパティを導入することで適切に表現できる。

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

³ ただし、HTML5 に対応したブラウザが必要。