



細見 格 NEC 情報・メディアプロセッシング研究所  
長野伸一 (株)東芝 研究開発センター  
岡部雅夫 東京電力(株) システム企画部

# 次世代の医薬品開発を支える知識流通

**新**薬の開発や安全性評価には、膨大な知識とデータが必要になる。そこにLinked Dataの有効性を見出そうとするワールドワイドな取り組みとして、W3CのLinking Open Drug Data (LODD)がある。LODDは、医薬品や臨床試験に関する情報を幅広く集めてLinked Data化し公開している。本稿では、LODDの狙いと現状を概観するとともに、副作用問題と創薬を対象として、公開されているLinked Dataの活用可能性を考察する。また、日本の医薬品業界における関連動向や課題についても述べる。

## 医薬品業界の2010年問題

### ●ジェネリックに対する期待と問題

昨今、新聞や雑誌、テレビなどで「ジェネリック」という言葉をよく耳にする。ジェネリックすなわちジェネリック医薬品とは、特許の期限が切れた医薬品(先発医薬品)と同じ主成分・製法を用いた後発医薬品の総称である。

世界中の大手製薬企業各社は、1990年代に現在の主力製品に関する特許を数多く出願しており、それらが2010年頃から相次ぎ期限切れを迎える。すると、他の製薬企業が同様の後発医薬品を製造・販売できるようになり、従来と同じ効果を期待できる医薬品がより低価格で入手可能になる。国としても、年々増大の一途を辿る医療関連の社会保障費負担が

少しでも軽減されるように、ジェネリックの普及促進を図っている。しかし一方で、メジャーな医薬品が特許で保護されず低価格化によって収益力も低下すると、新薬に膨大な研究開発費を投じている大手製薬企業にとっては死活問題となる。これは、医薬品業界における2010年問題と呼ばれている<sup>1)</sup>。

### ●創薬に必要な情報

患者の治療のために市場で販売できる医薬品を開発するには、多岐にわたる非常に多くの情報が必要になる。同じ病気を持つ患者であっても、その個人差によって薬の効き目はさまざまであり、誰にでも必ず効果のある薬というものはないといわれている。こうしたことから、医薬品の開発(創薬)には、病気、症状、患部とその状態、患者の特性、遺伝子や膨大な種類と構造を持つタンパク質、生体の反応経路(パスウェイ)などに関する非常に幅広い情報が必要になる。さらに、これらさまざまな情報の組合せに対して、新たな構造を持つ薬の臨床試験を行い、そこで得たデータを検証しなければならない。加えて、薬単体での有効性と安全性が確認できた後も、複数の異なる薬や特定の食材との間の組合せ(飲み合わせ)問題にも対処していかなければならない。

新薬やその候補に対し、国内で約18,000種類あるという医療医薬品、患者の状態などとの天文学的な

組合せについて、臨床試験を通じた問題の洗い出しを1組織で行うことは実質的に不可能といえる。特許切れに起因する2010年問題だけでなく、多くの患者にとって安全かつ効果的な新薬開発の問題は、すでに個々の医薬品メーカーの手に余る状況になっている。

#### 公開知識の共有による創薬効率化

創薬技術の高度化により、作用機序の分かりやすい疾患には完成度の高い薬が製品化されるとともに、未解決の疾患の難易度が高くなってきている。それに伴って、研究から製品化までを手がける垂直統合型のビジネスモデルはリスクとコストが加速的に増大し、分業化と囲い込みが進んでいる。図-1は、製薬産業における知識と製品の流通に関する構造を示している。大学や公的研究所は、政府機関や民間企業のファンドをもとに、疾患に対してターゲットとなるタンパク質や、効果のある化合物の発見に関する研究に取り組む。製薬企業は、自ら研究に取り組む一方で、大学・公的研究所や新興企業と提携することにより、新薬の技術を取り込んでいる。逆に大学・公的研究所や新興企業は、巨額のコストがかかる新薬候補の大規模な臨床試験を大手の製薬企業にゆだねている。このように、これら3者間には、新薬に関する技術とノウハウが取引される市場が形成されている。

しかし、このエコシステムには負の側面も見られる。本来、創薬には暗黙知が多く、すり合わせが欠かせない特性を持つ分野である。創薬の専門分野化、分業化が進みすぎてしまい、複数分野の知識を必要とする難病の研究が難しくなっている。そのため、複数分野の専門家が協力し、研究データを共有することにより、難病研究に取り組み始めている。そのような事例の1つとして、アルツハイマー病に関するプロジェクトを紹介する<sup>2)</sup>。

2004年に米国で、国立衛生研究所、食品医薬品局、創薬や医療画像に関連の企業、大学、非営利団体が集い、人間の脳の中からアルツハイマー病の進行を示す生物学的マーカーを探す共同研究プロジェクトADNI (Alzheimer's Disease Neuroimaging Initiative) を立ち上げた<sup>☆1</sup>。アルツハイマー病研究

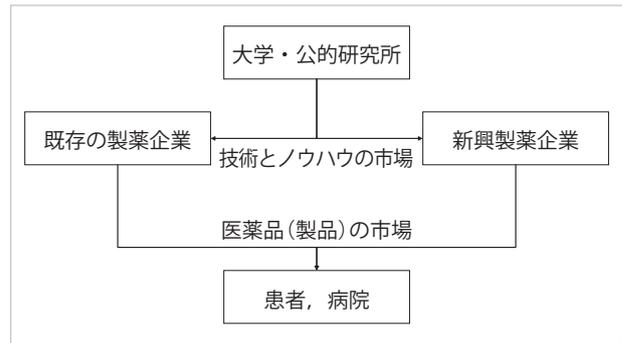


図-1 製薬産業のエコシステム

の問題は、複数分野の専門知識を必要とする難病であるにもかかわらず、多くの研究者が、彼らが所属する大学で彼ら自身の患者と手段とを用いて独自に研究を行っている点にあった。従来は共通のデータセットがないために研究成果の比較や検証が不可能であったことから、ADNIではどんな小さな発見であっても共有し、世界中のどこからでも計算機で利用できるようにすることを設立の目的とした。最近ではポジトロン放出断層撮影や脊髄液の試験などを利用したアルツハイマー病の早期診断法の研究が実を結びつつある。加えて、100件以上の臨床試験が進行中であり、アルツハイマー病の進行を遅らせたり止めたりする効果のある薬の発見が期待されている。同様の事例としては、パーキンソン病に関する共同研究プロジェクトがある。

ビジネスにおいても製薬会社が協力関係を模索し始めている。たとえば、米国のEli Lilly, Merck, Pfizerの3社は、アジアに患者が多い肺癌と胃癌の研究を進めてより良い治療法の開発を促進する目的で、2010年2月にAsian Cancer Research Groupを共同で設立した<sup>3)</sup>。また、特定の疾病に限定せず、世界中で散在しているライフサイエンス分野に関するさまざまなデータを統合的に利用できるように、データの収集や相互参照に取り組む事例も出てきている<sup>4)</sup>。海外ではW3CによるLinking Open Drug Data, Science CommonsによるNeuroCommonsプロジェクト, Bio2RDFプロジェクト, 国内ではライフサイエンス統合データベースセンターによる統合データベースプロジェクトなどがある。

☆1 <http://www.adni-info.org/>

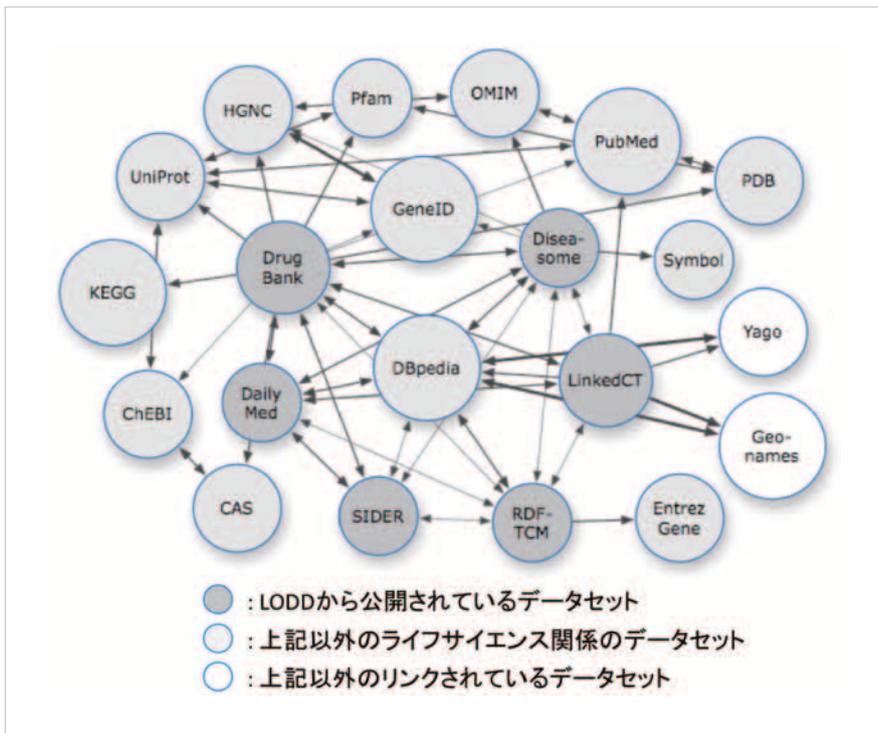


図-2 LODD クラウド(2009年8月6日現在)

めのタスクフォースとして、LODDが設立された。このLODDが設立された時期は、ちょうど、Linked Dataが注目を集め始め、いわゆるLODクラウドが急激に成長し始めた時期に符合する。

LODDは、すでにWeb上に、臨床試験情報、医薬品情報等、有用な公開情報があることを踏まえ、それらをLinked Dataとして相互にリンクさせることにより、さまざまな研究上やビジネス上の問題解決に役立つものにしようとするを意図している。LODDでは、それらの情報間に相互にリンクを張るために最新の研究成果を取り入れているのが特徴である。

図-2は、2009年8月時点で公開されているLODDに関連したLinked Dataのデータセットである。その後、さらにいくつかのデータセットが追加され(LODDから公開されているものは表-2

			リンク先						総計
			LODD + DBpedia(*)			左記以外			
			LODD	DBpedia	計	Bio2RDF	その他	計	
リンク元	臨床試験	LinkedCT	40,113	9,292	49,405	191,097	233,316	282,721	61,920
	薬	DailyMed	35,402		35,402				35,402
		DrugBank	21,921	1,522	23,443	38,554		38,554	61,997
	病気	Diseasome	8,502	6,124	14,626	14,048		14,048	28,674

(\*) DBpediaについては本特集第1編「Linked Dataの仕組み」の記事を参照

表-1 LODDにおける主なデータセットのリンク数

## Linking Open Drug Data (LODD)

### ● LODD とは

医薬品関係のLinked Dataのワールドワイドな事例として、W3CのLinking Open Drug Dataタスクフォース(以下、LODD)<sup>☆2</sup>での取り組みを紹介する。

W3Cでは、ヘルスケアやライフサイエンス分野でのセマンティックWebの活用を検討するためのグループとして、2005年9月からSemantic Web Health Care and Life Sciences Interest Group(以下、HCLSIG)が設置されているが、2008年10月に、その中に医薬品関係のLinked Dataを構築するた

参照)、その規模は少なくとも1,600万トリプル(トリプルは主語、述語、目的語の3つ組からなるRDFデータの単位)に上る。また、矢印で表されたデータ間リンクの数も、表-1のように大規模である。

### ● LODD の狙い

LODDの中核をなすデータセットの1つ、LinkedCT<sup>☆3</sup>の情報源であるClinicalTrials.gov<sup>☆4</sup>を例に、LODDの狙いとアプローチを考察する。

ClinicalTrials.govは、米国・保健社会福祉省下の国立医学図書館が運営している臨床試験情報を公

☆2 <http://www.w3.org/wiki/HCLSIG/LODD>

☆3 <http://linkedct.org/>

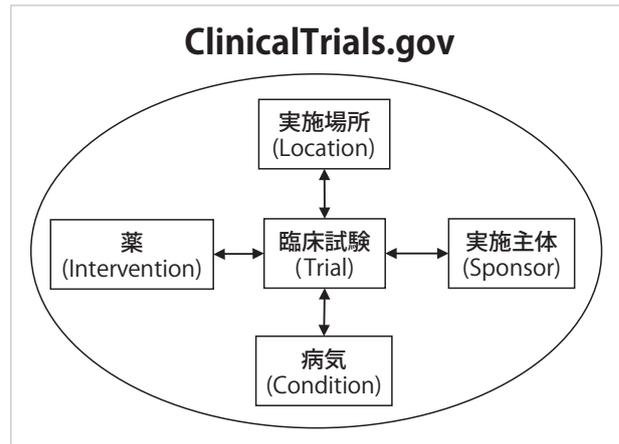
☆4 <http://clinicaltrials.gov/>

開するポータルである。臨床試験の透明性確保の観点から、米国では、1997年から生命を脅かす病気に対する臨床試験に関して登録が義務づけられており、ClinicalTrials.govはこれに呼応して開発され、臨床試験情報の登録・公開、被験者の募集の支援等のために用いられている。その後、世界保健機構(WHO)による臨床試験情報の公開義務付けの流れも受け、現在、世界174カ国から90,000件以上の臨床試験情報が登録・公開されている。

ClinicalTrials.govは、[図-3](#)に示す構造になっており、臨床試験を検索・ブラウズできるだけでなく、試験対象の病気、薬や、実施組織、実施地域からの検索やブラウズが可能になっている。したがって、このClinicalTrials.govの情報は、臨床試験の透明性確保という当初の目的を超えて、医薬品開発等のために幅広く活用できる可能性を秘めている。ただし、ClinicalTrials.govの情報を医薬品開発等の目的に利用しようとする、薬や病気そのものに関する情報が欠落している点が問題であった。

一方で、薬に関するWebポータルとしてはDrugBank<sup>☆5</sup>、DailyMed<sup>☆6</sup>等、病気に関してはThe human disease<sup>☆7</sup>等、臨床試験情報に限らずWeb上の公開情報は充実しつつある。しかしながら、ClinicalTrials.govが持つ薬や病気の情報は、[図-4](#)に示すようにこれらの情報とリンクされておらず、孤立状態になっている。

本来は、[図-5](#)のように、情報が臨床試験情報、薬情報、病気情報としてまとめ、それらが相互にリンクされているべきである。そのためには、対象について統一的な同一性の認識が存在し、かつ、その対象を指す統一的なIDが必要である。ただし、オープンでグローバルなWeb上の世界でこれらを実現することは現実的ではない。たとえば、アルツハイマー病は、ClinicalTrials.govでは“Alzheimer's disease”と表記されている一方で、The human diseaseでは“Alzheimer\_disease”と表記され、これらの名称が指示するアルツハイマー病という概念が厳密に同



[図-3](#) ClinicalTrials.govにおける臨床試験情報

一であるかどうかも定かではない。薬に関しても同様であり、たとえばClinicalTrials.govやDrugBankでは化合物としての一般名が用いられているが、DailyMedでは、商品化された薬を対象としているため商品名が用いられている。このほか、DrugBank、DailyMedは、それぞれ独自のID体系を導入している。

そこで、LODDはより現実的なアプローチをとっている。それぞれの公開情報は是として、それらの間にリンクを張ることにより、情報の孤立状態を解消しようとしている。具体的には、[図-5](#)における臨床試験と薬の関係、臨床試験と病気の関係は、ClinicalTrials.govが元々保持している関係を基本とし、[図-6](#)に示すように、ClinicalTrials.govの薬(intervention)とDrugBank、DailyMed、ClinicalTrials.govの病気(condition)とThe human diseaseそれぞれの間で、同一であることを表すowl:sameAsや参考情報であることを表すowl:seeAlsoのプロパティでリンクすることにより間接的に情報共有を実現している。

さらに、[図-5](#)における薬と病気の関係に関しては、DrugBank、DailyMedからThe human diseaseへは、possibleDiseaseTargetプロパティにより効果がある可能性のある病気に対するリンクを張り、逆にThe human diseaseからDrugBank、DailyMedへは、possibleDrugプロパティにより、効能を持つ可能性のある薬へのリンクを張っている。また、DrugBank(一般名の薬, genericDrug)とDailyMed(商品名の薬, brandedDrug)の間にもリンクを張っている。

☆5 <http://www.drugbank.ca/>

☆6 <http://dailymed.nlm.nih.gov/dailymed/>

☆7 [http://www.nd.edu/~alb/Publication06/145-HumanDisease\\_PNAS-14My07-Proc/Suppl/](http://www.nd.edu/~alb/Publication06/145-HumanDisease_PNAS-14My07-Proc/Suppl/)

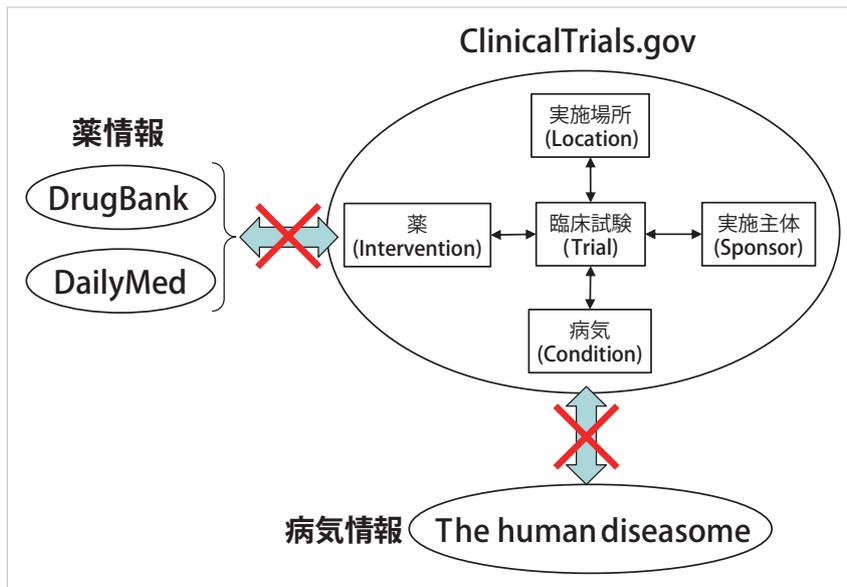


図-4 孤立状態にある ClinicalTrials.gov の情報

### ●LODDクラウド構築の実際

実際には、ClinicalTrials.gov 等の公開情報に直接リンクが張られているわけではない。情報源としての ClinicalTrials.gov, DrugBank, DailyMed, The human diseasesome は、それぞれ LinkedCT, DrugBank, DailyMed, Diseasesome と呼ばれるオリジナルのデータセットが RDF 形式の別のデータセットに変換され、それらの間に RDF によるリンクが付加され、Linked Data となっている。

動作環境としては、D2R サーバ<sup>☆8</sup>が利用され、必要な定義を行うことにより、関係データベース上で管理されている情報が自動的に Linked Data として公開される (D2R サーバについては本特集第1編「Linked Data の仕組み」の記事を参照)。情報源が別にある場合は、情報は D2R サーバ上の関係データベース用に変換される必要があり、また、必要な

☆8 <http://www4.wiwiw.fu-berlin.de/bizer/d2r-server/>

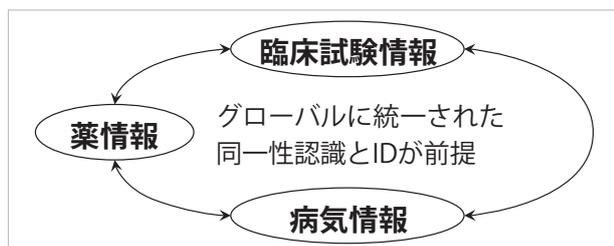


図-5 各種の医薬関連情報が結びついた望ましい姿

リンク情報もこの関係データベース上に付加されなければならない。たとえば、ClinicalTrials.gov から LinkedCT を構築するにあたっては、XML 形式を経て、関係データベースに変換するというステップが踏まれ、また、リンクの構築に関しては、関係データベース間のリンクを発見するシステムである Linkage Query Writer (LinQuer) が用いられ、文字列マッチングとオントロジーベース・マッチングによりリンクの発見が行われている<sup>5)</sup>。

文字列マッチングにおいては、文字列の類似性を評価することにより、たとえば、アルツハイマー病に関して“Alzheimer's disease” (LinkedCT) と“Alzheimer\_disease” (Diseasome) といった異表記だけでなく、“Adenocarcinoma of the Colon” と“Colon adenocarcinoma”といった表記ゆれまでを吸収している。ただし、“AIDS”が“Acquired Immunodeficiency Syndrome”の略記であること、“Klinefelter Syndrome”と“Hypogonadotropic hypogonadism”が同じ病気を表していることなどには対応できない。このようなケースに対応するため、さらにオントロジーベースのマッチングも行われている。具体的には、医療ドメインでの有名なオントロジーの1つである NCI シソーラスを用いて、これらの病名が同じ病気を表すものであるとの認識を行っている。

表-1 に LinkedCT, DailyMed, DrugBank,

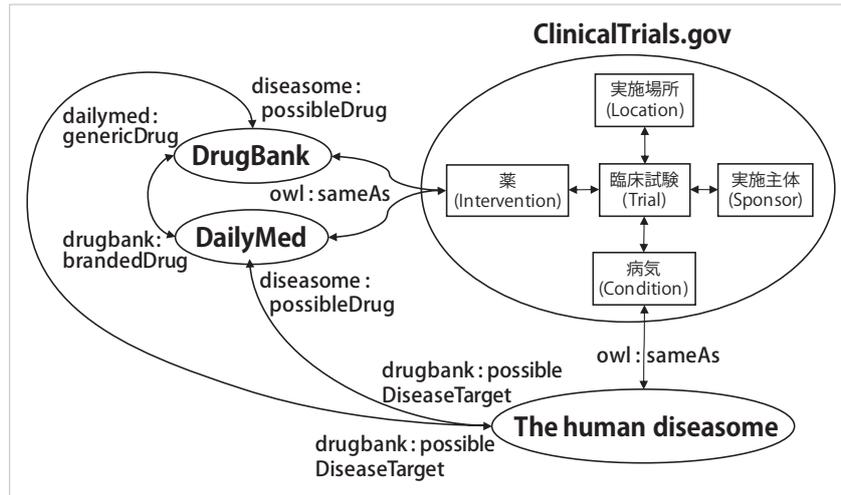


図-6 LODDにおける情報間リンクのアプローチ

Diseasome のリンク数の概略 (2009年4月現在) をまとめた<sup>6)</sup>。

LinkedCT では、実際には、DailyMed, DrugBank, Diseasome などの LODD のデータセットへのリンクよりも、LODD 側から外部の Linked Data へのリンクの方がはるかに多い。その中で、バイオ関係の公開情報を Linked Data 化するプロジェクトである Bio2RDF プロジェクト<sup>☆9</sup>のデータセットに対するリンクは、Bio2RDF's PubMed に対するものである。その情報源である PubMed<sup>☆10</sup>は、ClinicalTrials.gov と同じく米国・国立医学図書館が運営する医療文献に関する公開データベースであり、ClinicalTrials.gov の臨床試験情報は、関連文献としてこの PubMed へのリンク情報を保持している。このリンク情報をもとに、LinkedCT から Bio2RDF's PubMed へのリンクが生成できる。また、LODD から外部の Linked Data へのリンクには、PubMed 以外にも、臨床試験の実施地域から地理空間情報の Linked Data の1つである GeoNames<sup>☆11</sup>へのリンク、および情報源である ClinicalTrials.gov の該当 Web ページへのリンクがある。

DrugBank のリンクに関しても、半数以上が LODD 側から外部の Linked Data へのリンクであり、それらはすべて Bio2RDF のデータセットに対するものである。情報源側の DrugBank も、バイオ関係

の公開情報に対するリンク情報を多数保持しており、これらも情報源側のリンク情報を該当する Bio2RDF のデータセットへのリンクに容易に変換できる。

以上のことから、Linked Data としてのリンクの充実という意味では、元々の情報源間に共通の ID 情報を保持していない場合は、最新の手法によっても対応する情報間にリンク関係を発見することは必ずしも容易ではなく、元の情報源側に、統一的な ID ではなくとも、リンク先の ID を保持していることが、現実的には重要であることがうかがえる。

### ● LODD の現在

LODD で提供している Linked Data の最新の状況を主として LODD の Web サイトおよび各情報源の Web サイト<sup>☆2</sup>を元に、表-2 にまとめた。動作環境としては D2R サーバ以外のものが増えており、Linked Data として公開される情報量は着実に拡大してきている。これらのデータセットが実用に値するものであるためには、リンクを含め、情報が確実に更新され、情報の鮮度が維持されることが重要であるが、個々のデータセットの更新状況に関しては、必ずしもはかばかしいとはいえない。

たとえば、中核的なデータセットの1つである LinkedCT に関しては、情報源である ClinicalTrials.gov に登録されている臨床試験数は 95,547 件 (2010年9月15日現在) に増加しているが、LinkedCT では 61,920 件のままであり、2009 年前半に試験的に

☆9 <http://bio2rdf.org/>

☆10 <http://www.ncbi.nlm.nih.gov/pubmed>

☆11 <http://www.geonames.org/>



データセット	内容	情報ソース		データセット				
		名称	データ数	データ数	トリプル数	リンク数	データ更新	動作環境
LinkedCT	臨床試験	ClinicalTrials.gov	臨床試験：95,547	臨床試験：61,920	7,035,974	235,998	試験リリース	D2R Server
DrugBank	薬	DrugBank	薬：4,774	薬：4,772	765,936	59,661	定期的	D2R Server
DailyMed	薬	DailyMed	薬：10,553	薬：4,308	164,276	38,020	定期的	D2R Server
Diseasome	病気, 遺伝子	The human disease network	不明	病気 (+遺伝子等) : 4,213	91,182	31,065	2006年	D2R Server
SIDER	病気, 副作用	SIDER(Side Effect Resources)	薬：888 副作用：1,450	薬：924 副作用：1,737	192,515	20,555	2009年	D2R Server
STITCH	化学品, タンパク質	STITCH (Search Tool for Interactions of Chemicals)	小分子：74,000 タンパク質：2,500,000 有機体：630	化学物質：7,500,000 タンパク質：500,000 有機体：370	不明	不明	2009年7月	D2R Server
Medicare	高齢者向けの薬剤処方情報	Medicare.gov	不明	薬：6,819	44,500	4,959	不明	D2R Server
RDF-TCM	遺伝子, 病気, 薬, 成分	TCM GeneDlt	不明	遺伝子：945 生薬：848 成分：1,064 病気：553 作用：241	111,021	3,438	2009年8月(安定)	OpenLink Virtuoso Universal Server
ChEMBL	化学品, 分析(タンパク質, 有機物), 論文	不明	対象：8,054 化合物：600,625 論文：38,029等	不明	24M以下	不明	2010年1月	OpenLink Virtuoso Universal Server
RxNorm	薬	UMLS RxNorm	不明	薬および成分：165,806 用語：332,754	6,249,504	不明	2010年3月版 2010年5月更新	OpenLink Virtuoso Universal Server
WHO Global Health Observatory	感染症, 人口統計, 社会経済状況, 環境要因	WHO Global Health Observatory	項目：354,300	項目：22,384	152,000	不明	2010年9月更新	Triplify

表-2 LODD が提供するデータセット一覧 (<http://www.w3.org/wiki/HCLSIG/LODD/Data> の2010年10月のデータより)

リリースされたのち、新たなデータセットに対するリンクの付加や修正以外はあまり更新されていない。また、DrugBank と DailyMed に関しては定期的に更新されているとのことだが、その頻度は定かでない。

このように必ずしも更新が頻繁に行われていない理由としては、特に、LinkedCT における前述の例のようなリンク発見処理の複雑さがあるといえよう。Linked Data がリンク情報を含めて必要な更新が行われ、実用に耐えるデータの鮮度を維持していくためには、先にも述べたように、情報源側に最初からリンク先の ID に相当する明確なリンク情報が埋め込まれていることが現実的である。その上で、D2R サーバのように、情報源を自動的に Linked Data として公開できる環境を活用すれば、情報源から Linked Data に定期的に変換する必要はなくなり、Linked Data を実用に耐える鮮度に維持していくことも可能になるだろう。

実際、2010年1月に公開された ChEMBL<sup>☆12</sup> は、詳細は未確認だが、Bio2RDF プロジェクトの

☆12 <http://www.ebi.ac.uk/chembl/>

Linked Data へのリンク情報をまとめた形になっており、ChEMBL そのものが情報源のようである。今後はこのように最初から Linked Data として蓄積、公開されるケースが増えることが期待される。

## Linked Data の活用

### ●医療情報 Linked Data への期待

現在、LODD は前章で述べたような多数のプロジェクで公開されている医薬関連情報を Linked Data 化しており、その一部は定期的にデータを更新している。さらに、LODD とは別の Bio2RDF プロジェクトでも、40 のバイオ、遺伝子、医療関連のデータセット (約 23 億トリプル) を Linked Data 化し、SPARQL エンドポイントを公開している。こうした大量の医薬関連情報が公開され、まだ一部ではあるが定期的にも更新されていることで、創薬および製品化済み医薬品の利用に世界規模で貢献することが期待されている。



### ●副作用問題への活用可能性

典型的な活用法としては、特定のタイプの患者に対する医薬品の副作用(薬害を含む)情報の共有、特定の医薬品同士の飲み合わせ情報の共有などがある。患者のさまざまな特徴によるタイプの多さもさることながら、2万種類近く(一般医薬品も含めると4万種類以上)ある医薬品を複数同時に飲む場合の組合せは膨大な数になる。副作用被害は、国内では正確な情報がないものの、米国では患者数が年間約200万人、死亡者数が年間約10万人と報告されており、4番目に多い死因だという<sup>7)</sup>。

こうした副作用被害について可能な限り早くかつ詳しい情報を共有するために、LODDのようなオープンで標準的な形式に基づく世界規模の情報リポジトリが非常に重要と考えられている。米Eli Lilly社は、フェーズII(少数の軽度患者での臨床試験)～IV(市販後の副作用等の調査)のデータをClinicalTrials.govに2002年10月から登録しているとともに、独自にLillyTrials.comというサイト<sup>☆13</sup>でフェーズI(健常者に対する初期の臨床試験)の情報を2009年10月より公開している。その目的も、疾患や医薬品に関する臨床試験結果を参考情報として活用できるようにするためであり、実際の判断は医師にゆだねなければならないが、世界中の医師や患者に素早い気付きを与えることを目指している。

### ●創薬への活用可能性

一方、創薬への貢献に関しては次のような記事がある。「外国で新薬が発売されてから日本で発売されるまでに平均で4.7年かかる。米国(1.2年)の約4倍の期間だ。特に難病やがんなどで患者数の少ない治療薬は予想販売量や治験に参加できる日本人患者が少ないため、国内での開発が遅れがちだった」<sup>8)</sup>。

このように、日本などの一国内だけでは患者数が限られることから新薬の認定に時間がかかるため、世界中の臨床試験情報を活用して判断の迅速化を図ることも考えられる。しかし、こうした創薬の過程では非常に詳細な情報が必要となり、加えて患者のプライバシーにかかわる情報も多数含まれるため、

幅広い共有も困難である。

Linked Dataは、分野を横断して幅広い情報を緩やかに結び付け、さまざまな角度から気付きを与えると同時に素早く関連情報にアクセスできるようにすることを目指した技術であり、緻密な情報の共有や非常に厳密な情報間の照合といった目的に適しているとは言い難い。そうした目的には、まず対象領域における語彙や情報構造を標準化したオントロジーが必要となる。医薬品に関するそうしたオントロジーにはGene Ontology(GO)やOpen Biological and Biomedical Ontologiesなどがあり、特にGOはバイオインフォマティクス分野で最初の、かつ最も成功したオントロジーといわれるようになった<sup>9)</sup>。

しかし、こうしたオントロジーの構築と活用には、質と量の問題がある。すなわち、その分野の研究者が期待する正確さと詳細さを維持しつつ、幅広い表現やさまざまな関連領域の概念との関係を扱うことも求められる。質と量を兼ね備えたオントロジーの構築やメンテナンスには一流の人材と多大な労力が必要であり、これまでのLinked Dataのように急速な発展は難しい。

以上から、医薬関連情報のLinked Dataは、薬の副作用情報を世界で共有し素早い気付きを与えることで、リスクの高い飲み合わせの回避や、患者に合った薬の選択に寄与することが適しており、またそうした副作用問題の深刻さからも取り組むべき価値があると考えている。

### 日本の現状と今後に向けて

日本でClinicalTrials.govのように政府主導で医薬関連情報を扱っている機関としては、PMDA(独立法人医薬品医療機器総合機構)がある<sup>☆14</sup>。PMDAは、医薬品に関しては新医薬品や後発医薬品の審査のほか、薬事法で義務化されている医療機関や製薬会社からの副作用報告を整理・検証する役割を担っている。年間の報告件数は2009年までの過去5年で3.6万～4.7万件と増加傾向にある(2009年は減少)<sup>☆15</sup>が、検証には数年かかり、その情報は一般には非公開である。

☆14 <http://www.pmda.go.jp/>

☆15 <http://www.pmda.go.jp/operations/shonin/info/chikenkanren/fukusayou.html>

☆13 <http://lillytrials.com/>

一方、日本における治験・臨床研究登録機関である Japan Primary Registries Network (JPRN) が、世界保健機関(WHO)により WHO 指定の治験・臨床研究登録機関 (WHO Primary Registry) として 2008 年に認められた<sup>10)</sup>。JPRN の公開データを Linked Data 化していくことで、日本も医療系クラウドへ仲間入りし、よりワールドワイドな薬の効果と安全性の検証ができるようになることを期待したい。

なお、前章で述べたように、創薬では患者のプライバシーにかかわる情報を扱う必要がある。現在の治験や治療では、患者の個人情報が匿名化されており、さらに日本では、匿名化された情報から再び個人情報を特定できない連結不可能匿名化が実施されている。しかしこれでは、ある医薬品に重大な副作用が発見された場合も、その医薬品を適用した患者を特定できない。今後の創薬研究の発展には、個人情報の取り扱いに十分配慮した上で連結可能性を残せる個人情報保護技術の確立と、研究用途に限定した個人情報の二次利用のためのガイドライン整備が必要になるだろう。

### 参考文献

- 1) 佐藤健太郎：医薬品クライシス 78 兆円市場の激震，新潮社 (2010)。
- 2) Sharing of Data Leads to Progress on Alzheimer's, The New York Times (2010.8.8)。

- 3) Merck & Co., Inc. : Lilly, Merck, and Pfizer Join Forces to Accelerate Research and Improve Treatment of Lung and Gastric Cancers in Asia.  
<http://www.merck.com/newsroom/news-release-archive/research-and-development/home.html> (2010. 2. 23)
- 4) 中尾光輝, 片山俊明：分散データの統合とセマンティック Web, 情報処理, Vol.50, No.9(Sep. 2009)。
- 5) Oktie, H., et al. : LinkedCT : A Linked Data Space for Clinical Trials, CoRR (2009)。
- 6) Anja, J., et al. : Enabling Tailored Therapeutics with Linked Data, LODW (2009)。
- 7) 岡本 裕：9 割の病気は自分で治せる, 中経出版(2009)。
- 8) 「ドラッグ・ラグ」解消へ 国, 未承認薬など 109 品目開発要請, MSN 産経ニュース(2010. 5. 9).  
<http://sankei.jp.msn.com/life/body/100509/bdy1005092216000-n1.htm>
- 9) 榎屋啓志：生物学関連のオントロジーの最新動向, 人工知能学会誌, Vol.25, No.4, pp.485-492(2010)。
- 10) 厚労省：世界保健機関による日本の治験・臨床研究登録機関の認定について。  
<http://www.mhlw.go.jp/topics/2008/10/tp1017-1.html>

(平成 22 年 10 月 29 日受付)

### 細見 格 (正会員) i-hosomi@ay.jp.nec.com

日本電気 (株) 情報・メディアプロセッシング研究所主任研究員。自然言語処理・セマンティック技術の研究開発に従事。人工知能学会会員。

### 長野 伸一 (正会員) shinichi3.nagano@toshiba.co.jp

(株) 東芝研究開発センター ヒューマンセントリックラボラトリー主任研究員。博士 (工学)。2010 年より法政大学兼任講師。セマンティック Web, マルチエージェントシステムの研究開発に従事。電子情報通信学会, IEEE Computer Society 各会員。

### 岡部 雅夫 (正会員) okabe.masao@tepeco.co.jp

東京電力 (株) システム企画部スペシャリスト (ソフトウェア開発技法)。ISO/IEC 19763 Part3, Part10 エディタ。人工知能学会, 電子情報通信学会各会員。

