

平成 25 年度 厚生労働科学研究費補助金
(政策科学総合(統計情報総合)研究事業)
総括研究報告書

ICD11 オミックスサブ情報モデル(iCOs)の妥当性に関する実証研究

研究代表者 中谷 純(東北大学 大学院医学系研究科 教授)

研究要旨

本研究全体の目的は、2015年に改定が予定されているWHO-ICD11のコンテンツモデルに組み込むことのできるICD11オミックスサブ情報モデル(ICD11 Omics Sub information model: iCOS)の妥当性と性能に関する実証的検討を、東京医科歯科大学網羅的疾患分子病態データベース iCOD などの臨床オミックスデータベースの臨床オミックス実データ、GO, SNOMED-CT, NCK などのオントロジー実データを用いて行うことである。本年度においては、平成24年度に拡張を行ったiCOsモデルに基づき、まず多施設臨床オミックスデータベースを構成する上での課題の検討を行い、次にその解決方針に基づいて実データによるデータベース構築を行った。結果、SSMIX-2と親和性の高い、多施設間の臨床オミックスデータの共有・交換に資する中間ストレージが構成できることを確認した。また実データを用い共通参照レイヤーの情報粒度において、横断的な検索が可能であること、並びに各患者の臨床オミックスデータから、同一iCOsモデルのセマンティクスでWHO-ICD11, NCK等の外部疾患知識ベースの関連情報と対応づけられることを確認した。

A. 研究目的

様々な医学分野においてオミックス情報を利用することが必須となってきた背景において、日本から発信されたICD11オミックスサブ情報モデル(ICD11 Omics Sub information model: iCOs)の実用性を多角的に検証しWHOをはじめとした国際社会に示すことは、日本がオミックス医療という新分野で主導権をとりつつ、2015年に改訂が予定されているICD11のオミックス領域への可用性を高め国際社会に貢献することのできる重要な課題である。日本発信のiCOsは、やはり日本発信でISOにおいて正式に国際標準(IS)として認められたGSVML(Genomic Sequence Variation Markup Language)情報モデルを応用して作られたサブ情報モデルであり、世界に先駆けて、実オミックスデータを用いてその妥当性を検討することは、日本が率先して行うべき必須の課題であり、ICD11の2015年リリースを控え急務である。

そこで本研究ではWHO-ICD11改訂に資するため、ICD11のコンテンツモデルに組み込むことのできるICD11オミックスサブ情報モデル

(iCOs)の妥当性と性能に関する実証的検討を、東京医科歯科大学網羅的疾患分子病態データベース iCOD¹などの臨床オミックスデータベースの臨床オミックス実データ、NCKなどのオントロジー実データを用いて行うことを目的とする。本年度は平成24年度の成果に基づき、上記実データを用いたデータベース構築を行い、提案方式の妥当性並びに限界性能について検証を行った。

B. 研究方法

昨年度、拡張を行ったICD11オミックスサブ情報モデル(ICD11 Omics Subinformation model: iCOs)をデータベース構築のためのモデルとし、東京医科歯科大学の統合的臨床オミックスデータベース(Integrated Clinical Omics Database: iCOD)を検証用の実データとして用いた。iCOsモデルにはWHO-ICD11の特性記述モデル(Content Model)が包含され、ICD11コンテンツ自体をiCOsフォーマットで表現することが可能となっているが、現在のところICD11は改訂作業中でベータ版の段階であり、

¹ <http://www.tmd.ac.jp/mri/icod/index.html>

最終コンテンツの公開には至っていない。そこで、改訂の最新動向について厚生労働省 ICD11 国内内科 TAG 検討会に参加し、情報収集を行った。

iCOD は、統合データベースプロジェクトの一環として構築され、肝胆膵外科、大腸外科、及び口腔外科に入院した患者約 400 名を対象に、カルテから収集した病歴、診断、検査、外科的治療、内科的治療、予後の情報、及び看護師が聞き取った生活習慣情報と、病理標本よりレーザーダイセクションで切り出した患部における網羅的遺伝子発現情報を集積したものである。全体で 34 個のテーブルからなるリレーショナルデータベースとして構成されている。本研究では実データを iCOs モデルで格納しデータベース構築を行うが、目的はモデルの限界性能の検証であるため、各患者の症例の詳細内容を直接用いる訳ではない。そこで iCOD に対して連結不可能匿名化を施した上で、類似ダミーデータに置換した結果を借用し研究を行った。

次に iCOs モデルで実データを格納し、多施設の臨床オミックスデータベースを構築する上での課題を検討した。iCOs はデータの関係を記述した情報モデルでありデータベーススキーマではない。実際には患者数が膨大になった場合の格納方法、ICD11 や NCK といった静的な外部知識と動的に追加・変更される患者情報の切り分けの問題など、実データ格納上の検討課題があり、それら技術的要請に対する解決方針を定めた。

最後に上記解決方針に基づき、iCOD 実データを iCOs 形式の XML に変換し、データベースを構築した。また構築されたデータベースに対し、臨床オミックス研究で必要と考えられる検索の可否について検討を行った。

C . 研究結果

以下、本年度の主要な結果の概要について述べる。iCOs の詳細については、別添の図(K-Z)ならびに昨年度の報告を参照されたい。

C-1 臨床オミックスデータベースの格納における課題

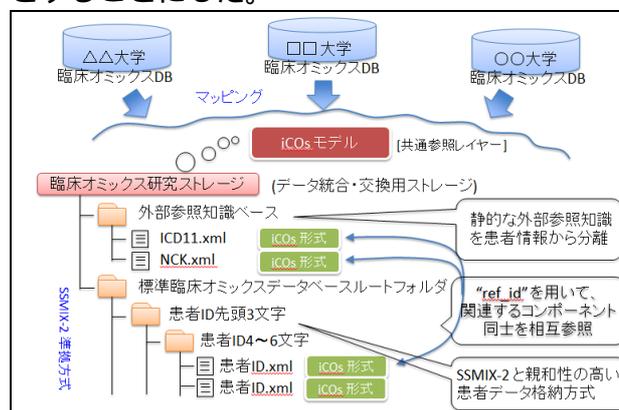
iCOs モデルは、(別添図 A,B) に示すように主に (1) WHO-ICD11 や疾患知識データベース NCK (Normalized Clinical Knowledge) といった外部参照医学知識、(2) 各患者のオミックス情報、(3) 各患者の臨床情報 の 3 つから構成されている。このうち、(2),(3) は(別添図 C) に示

すように <OML> という「ある時点の患者オミックス/臨床情報」と、<Time Course of Clinical Information> という「患者の OML 時系列と予後情報」によって記述されるようになっている。このように iCOs 自体は患者情報と外部知識リソースの双方を一元的に関連づけて記述するための情報モデルであるが、実際の臨床オミックスデータを格納するにはいくつかの技術的な要請による考慮事項が存在する。

- 前述の(1)は変更がほぼない静的なデータである。一方(2),(3)は患者数だけ存在し動的に追加・変更が行われるデータであるため、両者を分離して格納したい。
- 多施設の臨床オミックスデータの集約・交換のための中間ストレージとするためには、現在診療情報用の形式として利用が進められている SS-MIX2 のように「取り扱いが容易」な構造にしたい。
- 患者臨床オミックスデータの日々の追加や変更に伴うデータベース変更操作が簡便に行えるようにしたい。
- iCOs 全体は非常に多くのコンポーネントが入れ子構造で表現されており、深い階層のデータ構造であるが、計算機による検索上の利便並びに計算コストを考えると、主要なコンポーネントはなるべく上位の階層で現れるようにし、深い階層は避けたい。これはデータの更新をしやすくするためにも重要である

C-2 解決方針の設定

C-1 のような技術的課題を解決するために、まずデータベース全体は図 1 に示すような構造とすることにした。



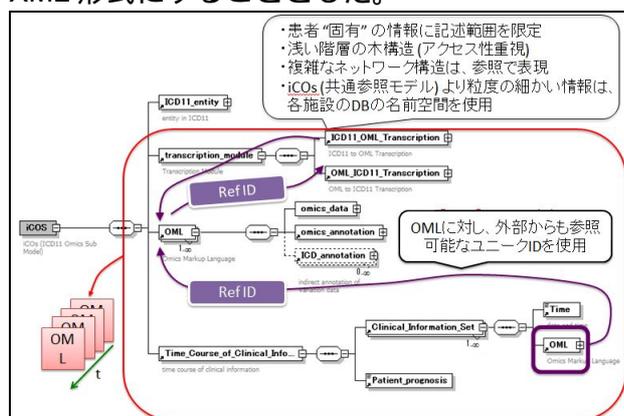
(図 1: データベース全体構造)

- 各施設の臨床オミックスデータベースは

iCOs モデル (共通参照レイヤー) で定められた情報粒度にて出力されるものとする。

- これらを蓄積・交換するための中間ストレージが、「臨床オミックス研究ストレージ」であり、「外部参照知識ベース」と「標準臨床オミックスデータベースルートフォルダ」からなる。
- 前者には、ICD11 や NCK といったコンテンツが iCOs 形式に則った XML にて格納される。後者は、SS-MIX2 と同様に、患者 ID の先頭 3 文字で区切られたフォルダ、さらにその中に患者 ID4~6 文字で細分化したフォルダを設置する。その中に各患者の臨床オミックスデータが、「患者 ID.xml」の名前で格納される。これも上と同様、iCOs 形式に則った XML とする。つまり、iCOs モデルの XML スキーマに基づいた 1 つの巨大な XML ファイルが同一形式の複数のファイルに分割されて格納されることになる。

一方、患者ごとのデータ構造は図 2 のような XML 形式にすることとした。



(図 2) 患者ごとのデータ構造

- 患者ごとの XML には、患者固有の情報のみを記述する。
- トップレベルは、「ある時点のオミックス/臨床情報」を記述する<OML>、「それらの時系列と予後」を記述する<Time Course of Clinical Information>、関連する ICD11 エンティティへのリンクである <Transcription Module> の 3 つとする。
- <OML> には外部からも参照可能な全データベース間で一意な ID を付与する。<Time Course of Clinical Information> や <Transcription Module> 内ではその ID を参照するだけで詳細は記述しない。これに

よりアクセス性を重視した、浅い階層の木構造とする。

- iCOs で定められた情報粒度より細かい情報については、各施設データベースの名前空間を用いた XML を入れ込むこととする。
- 分離された外部知識ベースとの接続についても (別添図 D) に示すように一意な参照 ID を用いることとする。

C-3 実データの格納

検証のため、C-2 で定めた解決方針に基づき、東京医科歯科大学統合的臨床オミックスデータベース(iCOD) の実データを元に、臨床オミックス研究ストレージを構築した。以下、別添図 E~J に具体的な XML での記述例を示す。

【別添図 E】は患者ごとのデータにおける上位構造である。OML の参照 ID 形式は、「施設 ID_患者 ID_日付(+同一日付内 ID)」とし、全データベース中で一意となっている。これにより、ICD11 の疾患記述から、関連する実患者オミックス情報を指定する事ができる。個々の OML はある時点での (1) オミックスデータ、(2) オミックスアノテーション情報、(3) 臨床情報、から構成されている。患者受診によりデータが追加される場合も、<OML> タグを追加し、<Time Course of Clinical Information> 内に日付と参照 ID を追加するだけで容易に変更可能である。

【別添図 F】は、OML 内の <Omics Data> 部分を示す。<Transcription Type>、<map>、<ass_gene_name> などの具体的な Omics データが格納されている。今回の iCOD を用いた検証では、iCOs(共通参照レイヤー) で規定された情報粒度のタグのみで、全ての情報が記述可能であった。

【別添図 G】は、上述の<Omics Data> に対するアノテーション情報である。ここでは、共通参照レイヤーよりも粒度の細かい情報を格納しなければならず、iCOD の名前空間を用いた小さな XML を各項目に格納している。

(例: <iCOD:分子情報:検体部位> など)。また、外部の図やテーブルは、本データベースに直接格納することが不適当と考えられたため、外部ファイルへの参照という形で記述した。

【別添図 H】は、患者の臨床情報を格納する

<ICD Annotation> 部である。患者の基本情報や、ICD11 への対応情報、外部疾患知識 NCK 内の該当項目へのポイント、具体的な<Clinical Observation> などが格納されている。

【別添図 I、J】は、<Clinical observation> 内で記述される臨床情報から、例として Life History と 画像検査情報 の記述例を示す。

Life History (図 I) については、共通参照レイヤーで詳細項目が定められていないため、

- <iCOD:運動>しない</iCOD:運動>
- <iCOD:喫煙歴>10年</iCOD:喫煙歴>

のように各 DB の名前空間を用いて記述されている。

画像検査情報 (図 J) では、CT 所見や内視鏡検査結果の記述がされているが、画像ファイルについては直接格納するのが不相当と考えられたため、外部画像ファイルへの参照ポイントが格納されている。

C-4 本方式で実現された内容

実データを用いた C-3 の検証結果により、本研究の方式にて下記の 3 つの特徴を持った臨床オミックス研究ストレージが構築できることが確認できた。

(1) 臨床オミックスデータの共有・交換

他施設間の中間ストレージとして機能し、SSMIX と親和性が高く、更新や変更などの操作が簡便な形式でのデータベース構成を行った。

(2) 共通粒度における横断的検索

昨年度成果により分野専門家の合意で定められた共通参照レイヤーの情報粒度においては、iCOs で規定されたタグとその間の参照関係を用いて、横断的な検索が可能であった。

(例): <ass_gene_symbol>, 各種 <clinical observation> 項目を用い、「特定の遺伝子“AURKB”を持ち、” ”を発症した対象者」といった検索

(例): <chromosome_number>, <position> を用い、「同じ染色体・位置“17p13.1”に遺伝子を持つ対象者」といった検索

(3) 外部知識リソースとのデータの関連付け

特定の患者症例と、その疾患についての一般的な医学知識 (ICD11 の特性記述、NCK の疾患知識) が、同一 iCOs モデルのセマンティクスで対応づけられている。

一方、本方式では限界と思われた事項については、次の考察の中で述べる。

D . 考察

近年のオミックス医学の発達に伴い、臨床情報とオミックス情報を統合してデータマイニング手法により解析し、種々の医学的知見の発見に応用しようとする研究が盛んになりつつある。本研究で検証に用いた東京医科歯科大学の iCOD データベースもその一環であるが、今後も様々な施設でデータベースの構築が行われると考えられる。従って、将来的にそのような臨床オミックスデータを多施設間で統合的に解析するための方法論の確立は極めて重要である。本研究では、まず ICD11, NCK といった外部参照知識ベースと動的に追加・変更される患者臨床データ(iCOD) を分離し、SS-MIX2 と同様のデータ構造を採用した。これによりデータの追加や変更に伴う操作が簡便で取り扱い容易な、多施設間の情報共通・交換のための臨床オミックスデータ中間ストレージを実現しており、これが提案方式の大きな特徴の一つである。

また、外部参照知識ベースと各患者臨床データの全てが、同一の iCOs モデルに基づいて記述されており、各タグの参照 ID を用いて互いに情報を辿ることができる。これにより、特定の患者症例から、ICD11 の疾患特性記述を参照したり、逆に ICD11 の疾患記述から、その疾患に関連する患者のオミックスデータを取得したりすることが可能である。また「特定の遺伝子を持ち ” ”を発症した患者」といった横断的な検索も可能となっている。これらは多施設データベースの共通参照レイヤーとして、iCOs の情報粒度を統一的に用いた利点であり、これもまた本研究の大きな特徴となっている。

一方で、“共通参照レイヤー” よりも細かい情報粒度を用いた検索クエリに実現に向けては限界があることも判明した。例えば、「ある遺伝子の発現量が 2.0 以上で、かつ肝硬変の症状を呈した肝疾患の患者一覧」を取得する検索は、現在行うことができない。これは、必ずしも全てのデータベースが全遺伝子に対する発現量の情報を持っているわけではないため、iCOs(共通参照レイヤー)では「発現量」の記述項目を採用しなかったためである。同様に、各データベースでの記述項目が大きく異なると考えられるため iCOs では詳細が定められていない <Life History> 等の項目においても、現在

では各施設データベースの固有の名前空間を用いて、<iCOD:喫煙歴> のような形式でデータが格納されているため、多施設データベースにまたがった横断的検索が行えない。これは本研究のような2段階モデルを採用している電子カルテ情報交換のための規格 CEN13606 などでも見られる問題であるが、解決は容易ではない。CEN13606 では、共通参照レイヤーよりも細かい粒度の情報項目は Archetype という形式で記述し、ユーザーはリポジトリに蓄積された Archetype から利用したいものを選択するという方式を採用しているが、異なる Archetype の項目が意味的に同一であるかどうかをシステムティックに判定するための仕組みを提供していないため、問題を解決できていない。

これに対する将来的な1つの解決策としては、RDF データベースの利用が考えられる。RDF 自体は、データを <主語, 述語, 目的語> の3つ組で表現しグラフ構造としてデータベースを記述するための方式であるが、RDF の上位レイヤーである OWL を用いれば、<OWL:SameAs> というリンクで、エンティティの同一性を記述することができる。すなわち、各施設間でバラバラな詳細項目についても、例えば<iCOD:喫煙歴>と<XXDB:SmokingStatus>を<OWL:SameAs> で連結することで、後から両者の項目が同一であることが記述できる。将来的には、現在の中間ストレージ内のデータを全て RDF データベースにエクスポートし、必要に応じて項目の同一性を OWL レイヤーでメタ情報として記述し、SPARQL クエリを用いて横断的検索を行う、といった手法が考えられ、今後検討を行う予定である。

E . 結論

本年度は、昨年度に拡張を行った ICD11 オミックスサブ情報モデル(ICD11 Omics Sub information model: iCOs) に基づき、東京医科歯科大学統合的臨床オミックスデータベース iCOD の実データを格納する際の課題を明らかにし、解決方針に従って実データベースを構築した。結果、(1) SS-MIX2 と親和性の高いデータベース全体構造の導入、(2) 外部知識と患者臨床情報の分離、(3) Reference-ID を用いた複雑な参照構造の表現、(4) 共通参照レイヤーと各 DB 固有の名前空間の分離、といった工夫により、実データを問題なく格納できることを確認した。本研究の方式により構築された臨床オミックス研究データベースは、多施設間の情報

共有・交換のための中間ストレージとして機能し、共通参照レイヤーの情報粒度で横断的な検索が可能である。また個々の患者臨床データの情報項目は、iCOs モデルを通じて対応する ICD11 エンティティと対応づけられており、ICD11 のコンテンツモデルで記述された疾患特性 (一般的医学知識) と相互に参照することが可能である。

現在 ICD11 の改訂が進められているが、本研究で開発した iCOs モデルによる臨床オミックス研究データベース構築方式は、ICD11 のオミックス領域への可用性を高め国際社会に貢献するのみならず、今後ますます発展していくと見込まれる臨床オミックス研究の情報学的基盤として重要な役割を果たすと期待される。今後は、より細かい情報粒度での検索クエリの実現のため、RDF データベース上での SPARQL クエリの利用と OWL レイヤーでのメタ情報記述の組み込みを検討する予定である。

F . 健康危険情報 なし

G . 研究発表

1. 論文発表

- 1-1. S Ogishima, S Mizuno, M Kikuchi, A Miyashita, R Kuwano, H Tanaka and J Nakaya.2013. A Map of Alzheimer's Disease-Signaling Pathways: A Hope for Drug Target Discovery. Clinical Pharmacology & Therapeutics. 93,(20 February 2013) | doi:10.1038/clpt.2013.37. 399-401

2. 学会発表

- 2-1. 医療情報の標準化に関する戦略 中谷 純
厚生労働省 ICD 室 東京、2013.5.14
- 2-2 . これからの情報通信分野における取り組むべき課題
中谷 純 自由民主党本部/情報通信戦略調査会 東京、2013.5.22
- 2-3 . 未来型医療において必要な医療情報基盤 中谷 純
第 98 回東北医学会総会/教授就任記念講演会 仙台、2013.5.24
- 2-4 . みやぎ医療福祉情報連携基盤について 中谷 純
復興推進委員会 仙台、2013.5.25
- 2-5 . MMWIN について 中谷 純 総務省視察 仙台、2013.5.31
- 2-6 . 未来型医療に必要な医療情報基盤 中谷 純
老年医学会北海道地方会 札幌、2013.6.8
- 2-7 . 東北メディカル・メガバンク・MMWIN 統一的視点による医療情報基盤構築事業 先にあるべきもの
中谷 純 内閣官房 東京、2013.6.13
- 2-8 . 医療情報から見た次世代医療：ゲノム医療から先制医療へ
中谷 純 日本脳ドック学会総会 仙台、2013.6.21
- 2-9 . みやぎ医療介護福祉情報ネットワーク (MMWIN) について
中谷 純 地域医療福祉情報連携協議会第 5 回シンポジウム 東京、2013.6.28
- 2-10 . みやぎ医療福祉情報連携基盤 (MMWIN) 中谷 純
モダンホスピタルショー 2013 東京、2013.7.17
- 2-11 . 医療情報の国際標準化について 中谷 純

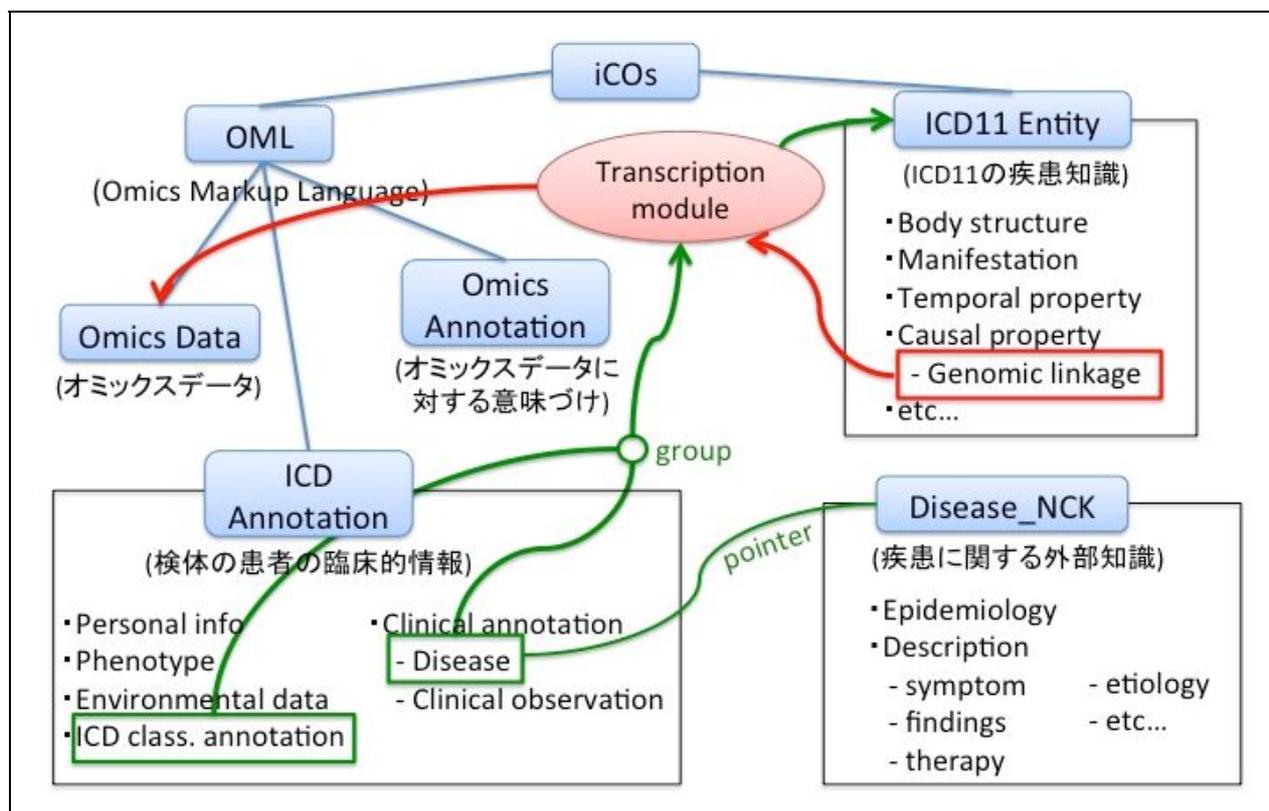
宮城県診療情報管理研究会 仙台、2013.7.27

- 2-12 . MMWIN について 中谷 純 総務省視察 仙台、2013.7.29
- 2-13 . Jun Nakaya. Essential Medical ITC Infrastructure in Future Medicine. Medinfo2013.
- 2-14 . みやぎ医療福祉情報連携基盤 (MMWIN) 中谷 純 MMWIN キックオフ祝賀会 仙台、2013.9.6
- 2-15 . 東北での取り組みのご紹介 中谷 純 東京大学公共政策大学院 リスクマネジメント講座 シンポジウム 東京、2013.9.18
- 2-16 . MMWIN による地域医療介護福祉連携 中谷 純 第 23 回日本医療薬学会年会講演 仙台、2013.9.21
- 2-17 . MMWIN 「システムの稼働状況について」「施設の参加状況と課題について」 中谷 純 平成 25 年度厚生労働科学研究費補助金 (地域医療基盤開発推進研究事業) 「被災地における地域医療情報連携体制のあり方に関する研究」 第 1 回班会議 仙台、2013.9.27
- 2-18 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 仙台医療圏説明会 (黒川郡医師会) 宮城、2013.10.10
- 2-19 . 宮城における地域医療介護情報連携 中谷 純 第 40 回日本産婦人科医会学術集会 仙台、2013.10.13
- 2-20 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 仙台医療圏説明会 (仙台市医師会 病院・福祉施設) 仙台、2013.10.16
- 2-21 . 未来の温泉医学 中谷 純 日本温泉気候物理医学会 東北地区研修会 宮城、2013.10.19
- 2-22 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 仙台医療圏説明会 (名取・岩沼・亶理医師会) 宮城、2013.10.21
- 2-23 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 仙台医療圏説明会 (仙台市医師会 診療所・調剤薬局) 仙台、2013.10.24
- 2-24 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 仙台医療圏説明会 (宮城県塩釜医師会) 宮城、2013.10.31
- 2-25 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 説明会 (ひかり薬局) 仙台、2013.11.6
- 2-26 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 説明会 (東北薬科大学病院) 仙台、2013.11.8
- 2-27 . MMWIN における慢性期医療への取り組み 中谷 純 第 21 回日本慢性期医療学会 ランチョンセミナー 東京、2013.11.15
- 2-28 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 説明会 (登米市) 宮城、2013.11.18
- 2-29 . 広域保健医療・連携医療支援_災害時/BCP 第 33 回医療情報学連合大会 座長 酒巻 哲夫、中谷 純 東京、2013.11.21
- 2-30 . みやぎ医療福祉情報連携 NW について 中谷 純 第 33 回医療情報学連合大会 企業共催セッション 東京、2013.11.23
- 2-31 . 仙台圏地域医療情報連携基盤構築事業説明 中谷 純 MMWIN 説明会 (宮城県立こども病院) 仙台、2013.11.29
- 2-32 . みやぎ医療介護福祉情報ネットワーク (MMWIN) について 中谷 純 宮城県放射線技師会第 22 回総合学術大会シンポジウム 仙台、2013.11.30
- 2-33 . 地域医療情報ネットワークについて 中谷 純 院内がん登録実務者養成コースセミナー 仙台、2013.12.10
- 2-34 . みやぎにおける地域医療情報連携の構築 中谷 純 プラチナ構想ネットワーク講演 東京、2013.12.13
- 2-35 . Tatsukawa A, Shinohara EY, Imai T, Kawazoe Y, Ohe K. An analysis of the openEHR archetype semantics based on a typed lambda theory. Stud Health Technol Inform. 2013;192:990. (poster) openEHR archetype semantics based on a typed lambda theory. Stud Health Technol

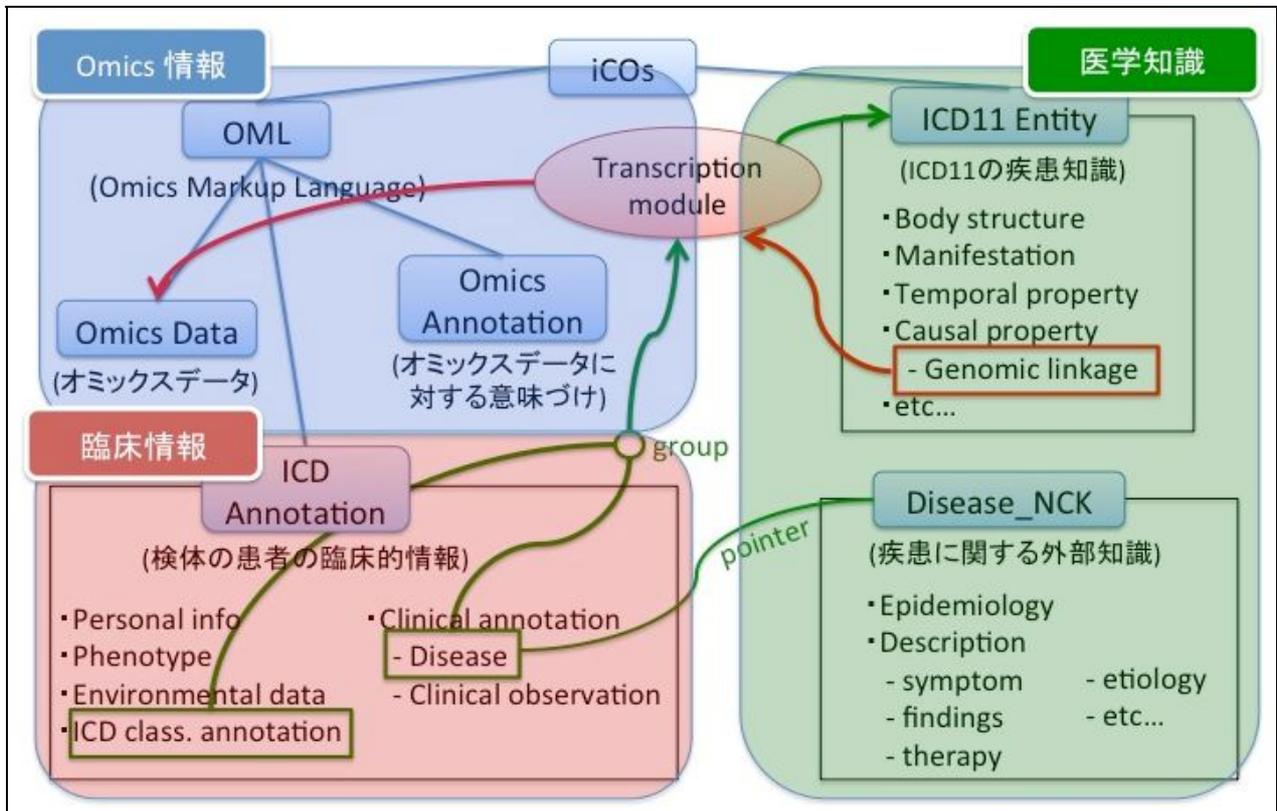
Inform. 2013;192:990. (poster)

H . 知的財産権の出願・登録状況
なし

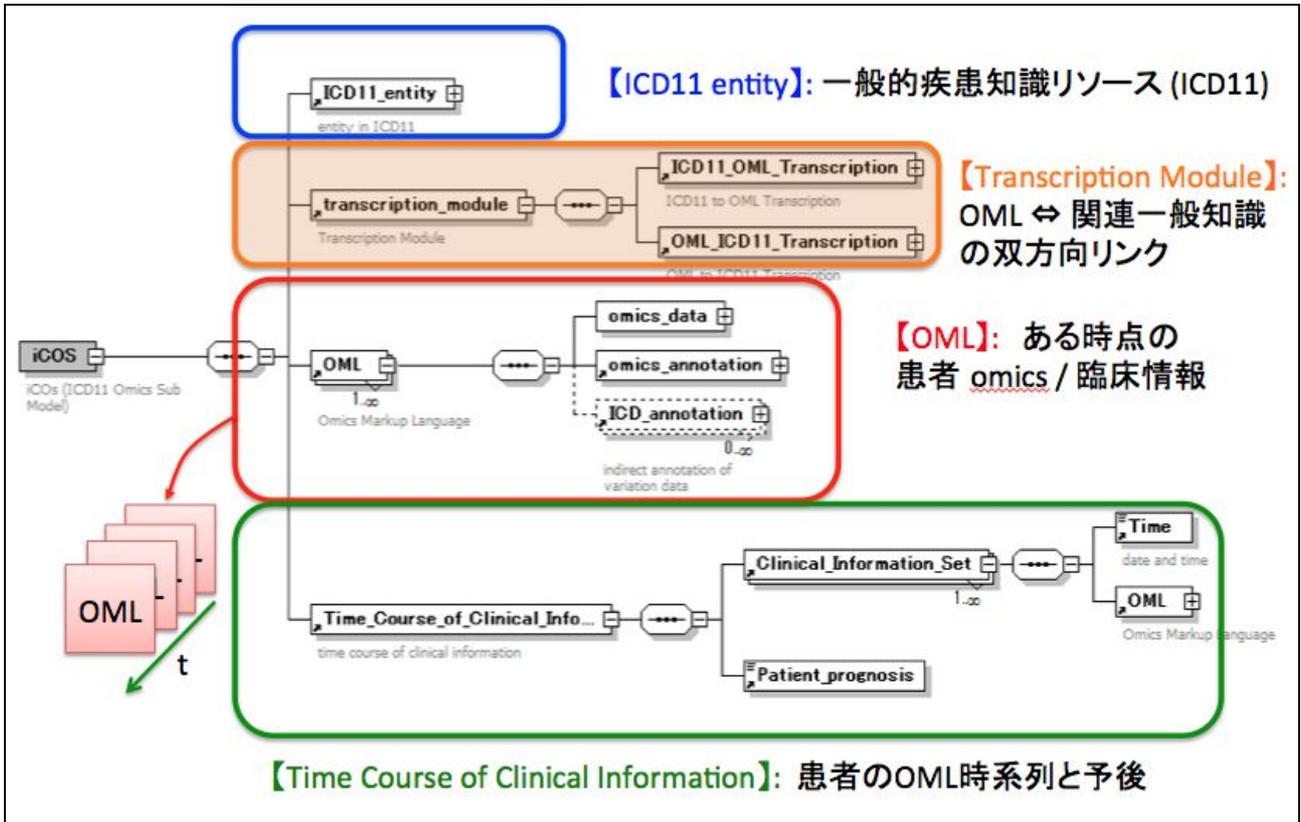
(別添資料)



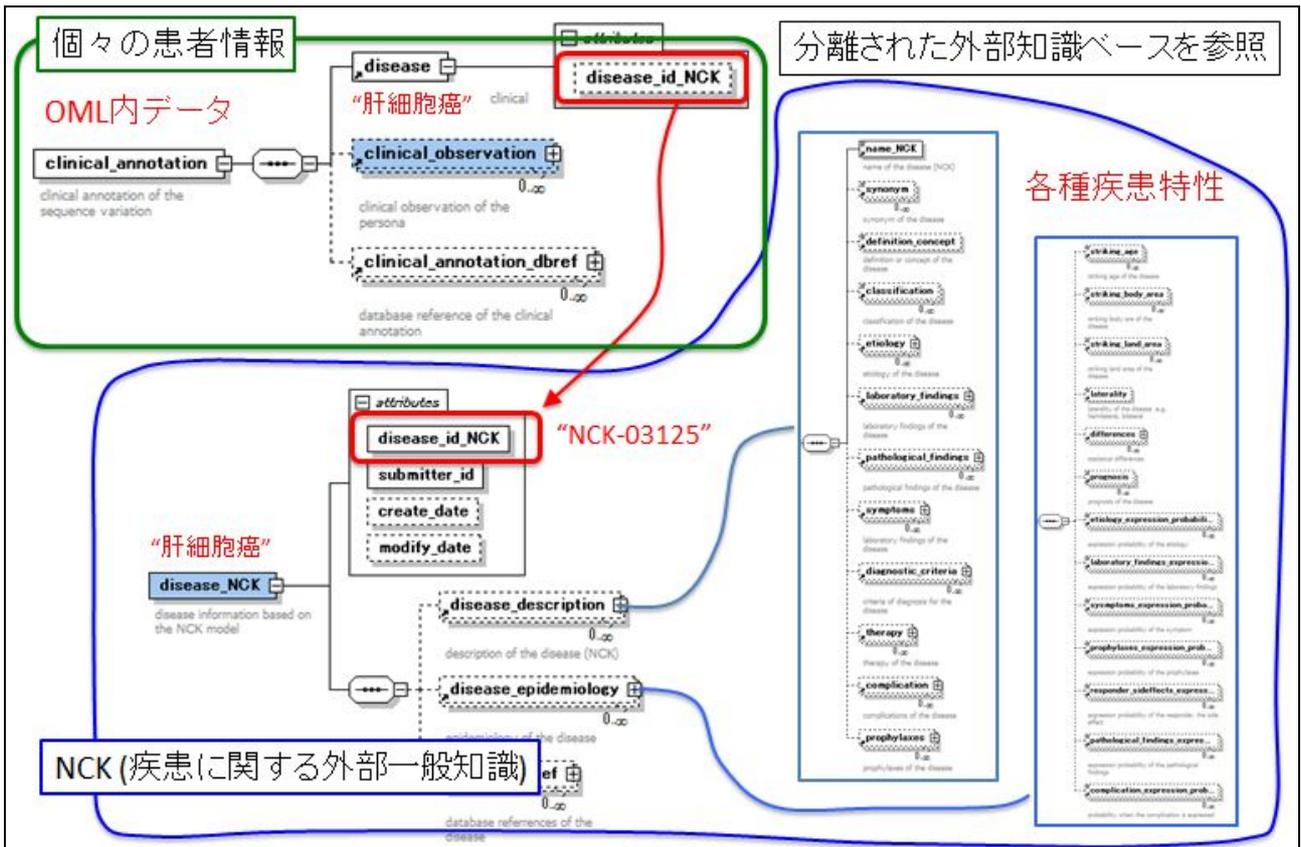
(図 A . iCOs モデル概要)



(図 B . iCOs モデルにおける記述情報種別)



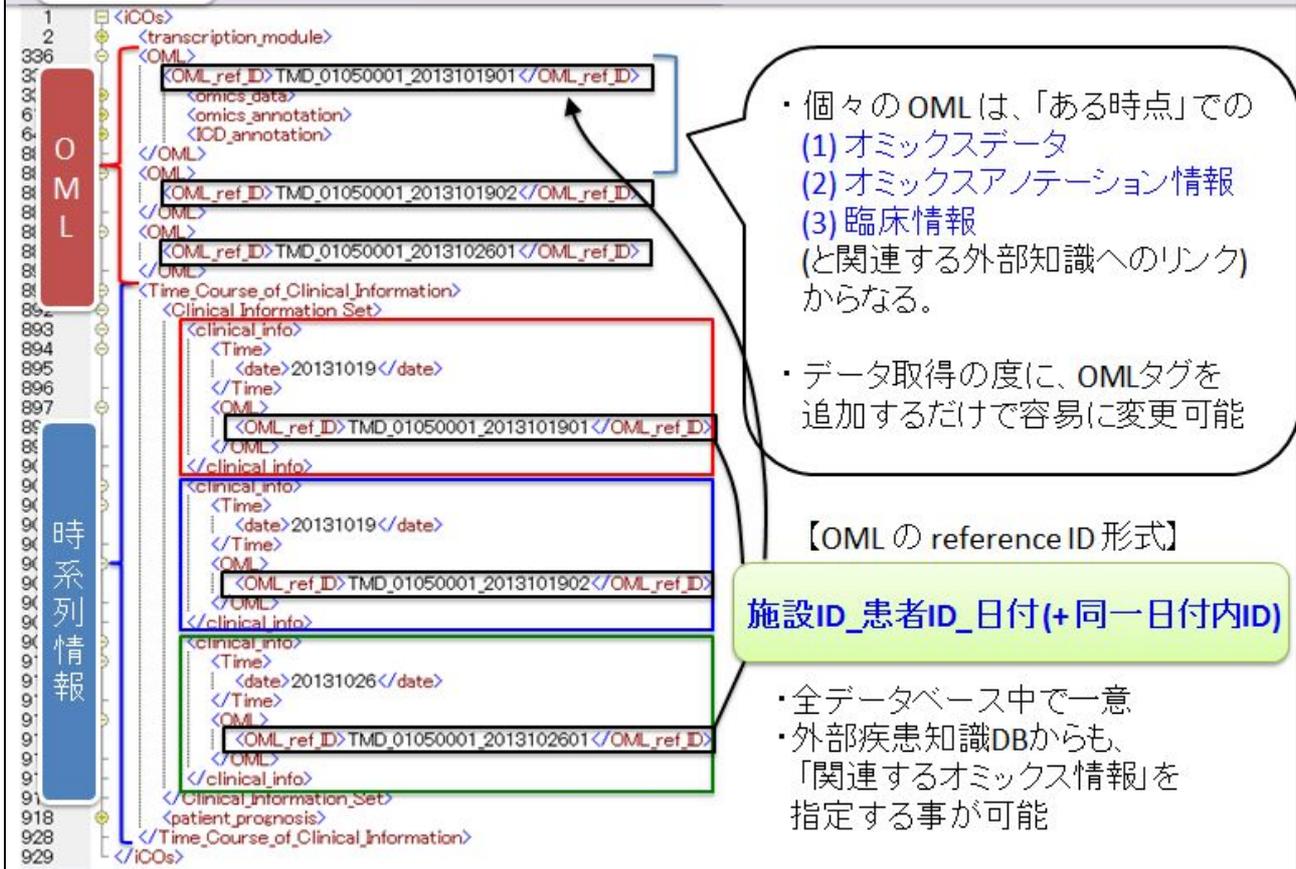
(図 C: iCOS XML スキーマトップ階層)



(図 D: 外部知識ベースとの接続)

具体例

患者ごとのデータ:上位構造 (OML reference ID)



(図 E: 患者ごとのデータ: 上位構造)

具体例

OML内のデータ構造 (1) Omics Data

```

17 <omics_data>
18 <variation_data>
19 </variation_data>
20 <transcription_data>
21 <transcription_data>
22 <transcription_type>mRNA</transcription_type>
23 <transcription_location>
24 <chromosome_number>17</chromosome_number>
25 <position>p13.1</position>
26 <map>Chromosome: 17; NC_000017.10 (8106049..8113944, complement)</map>
27 <orientation>/>
28 </transcription_location>
29 </transcription_data>
30 <ass_gene>
31 <ass_gene_name>aurora kinase B</ass_gene_name>
32 <ass_gene_structure/>
33 <aminoacid_substitution/>
34 <codon_substitution/>
35 <codon_position/>
36 <ass_gene_symbol>AURKB</ass_gene_symbol>
37 <ass_gene_alias>serine/threonine kinase 12, STK12, Aurora B, AurB, AIM-1, ARK2, STK5</ass_gene_alias>
38 <ass_gene_product>AURKB</ass_gene_product>
39 <ass_gene_evidence_type/>
40 <changed_motif/>
41 <changed_motif_name/>
42 <changed_splice_site/>
43 <splice_variant_number>
44 <refSeq_number>NM_001256634.1, NP_001243763.1</refSeq_number>
45 </splice_variant_number>
46 <ass_gene_dbref>
47 <database_name>NCBI Gene</database_name>
48 <database_id/>
49 <link_url>http://www.ncbi.nlm.nih.gov/gene/</link_url>
50 <database_attr/>
51 <data_id>9212</data_id>
52 <data_attr>Afymetrixx社のプロブID: 209464_at, 239219_at</data_attr>
53 <version>26-Jan-2014</version>
54 </ass_gene_dbref>
55 </ass_gene>
56 <location_dbref>
57 </location_dbref>
58 <transcription_location>
59 <transcription_att>
60 <source>
61 <source_release_date>2008/7/3</source_release_date>
62 <source_modify_date>2009/12/22</source_modify_date>
63 <source_raw_data>Not Available</source_raw_data>
64 <source_dbref>
65 <database_name>iCOD: Integrated Clinical Omics Database</database_name>
66 <database_id/>
67 <link_url>http://omics.tmd.ac.jp/</link_url>
68 <database_attr/> iCODは、医療への応用に向けたゲノム、プロテオームなどの分子情報の成果をこれまでの臨床・病理情報や環境情報と関連づけ、
69 新たな臨床医学として体系化するための、データベースを目指しています。2005年より収集したがん症例について、臨床・病理情報、環境情報および分子情報を利用し、
70 研究デザインに沿った統計解析、症例ごとの病態像について固相レベルから、組織レベル、細胞レベルに至るまで、横断的に視覚化するための情報を統合しています。</database_attr>
71 <data_id>01050001</data_id>
72 <data_attr>肝細胞癌140例、大腸癌106例、膵臓癌75例、食道癌91例のデータがあります。</data_attr>
73 <version>2009/12/22</version>
74 </source_dbref>
75 </source>
76 </transcription_att>
77 </transcription_dbref>
78 </transcription_data>
79 <proteomics_data>
80 <signalomics_data>
81 <metabolomics_data>
82 <organomics_data>
83 <other_omics_data>
84 </omics_data>

```

<Transcription_type> mRNA
 <map> Chromosome:17;NC_000017.10
 <ass_gene_name> aurora kinase B
 <ass_gene_symbol> AURKB
 <refSeq_number> NM_001256634.1, NP_001243763.1
 <database_name> iCOD

iCOs (共通参照レイヤー) で規定されたタグのみで
 上記の情報は記述可能

(図 F: OML 内のデータ構造 (1) Omics Data)

具体例

OML内のデータ構造 (2) Omics Annotation

```

607 <omics_annotation>
608 <transcription_experiment_ar
609 <differentially_expressed_ar
610 <new_sample_information>
611 <icod.bunnsi.jyouhou.kenntai.saisyu.syudann>4期</icod.bunnsi.jyouhou.kenntai.saisyu.syudann>
612 <icod.bunnsi.jyouhou.kenntai.bui>癌部</icod.bunnsi.jyouhou.kenntai.bui>
613 <icod.bunnsi.jyouhou.sample.comment>無</icod.bunnsi.jyouhou.sample.comment>
614 <icod.bunnsi.jyouhou.jyouonn.houti.jikan>15.52</icod.bunnsi.jyouhou.jyouonn.houti.jikan>
615 <icod.bunnsi.jyouhou.yondo.horei.jikan>0.28</icod.bunnsi.jyouhou.yondo.horei.jikan>
616 </new_sample_information>
617 <array_information>
618 <icod.bunnsi.jyouhou.jikken.syuhou>DNAマイクロレイ</icod.bunnsi.jyouhou.jikken.syuhou>
619 <icod.bunnsi.jyouhou.chip.no.syurui>GeneChip(U133plus2.0)</icod.bunnsi.jyouhou.chip.no.syurui>
620
621
622 <icod.bunnsi.jyouhou.kaiseki.data.suu>100</icod.bunnsi.jyouhou.kaiseki.data.suu>
623 <icod.bunnsi.jyouhou.algorith>RMA</icod.bunnsi.jyouhou.algorith>
624 <icod.bunnsi.jyouhou.filtering>なし</icod.bunnsi.jyouhou.filtering>
625 <icod.bunnsi.jyouhou.kennte>門脈侵襲または肝静脈侵襲(vp/vv) Wilcoxonの順位検定</icod.bunnsi.jyouhou.kennte>
626 <icod.bunnsi.jyouhou.idensti.senntaku>P-valueの小さいものから100プローブセット</icod.bunnsi.jyouhou.idensti.senntaku>
627 <icod.bunnsi.jyouhou.clustering>Z変換を行った後、ユークリッド距離、完全連結法で実行</icod.bunnsi.jyouhou.clustering>
628 <icod.bunnsi.jyouhou.heatmap>figure_h000001</icod.bunnsi.jyouhou.heatmap>
629 <icod.bunnsi.jyouhou.pathwaymap>figure_p00001</icod.bunnsi.jyouhou.pathwaymap>
630 </DEG_statistical_analysis>
631 <differentially_expressed_gene>
632 <icod.bunnsi.jyouhou.gene.list>table_x00001</icod.bunnsi.jyouhou.gene.list>
633 </differentially_expressed_gene>
634 <differentially_expressed_gene_DEG_anaylsis>
635 </transcription_experiment_analysis>
636 </omics_annotation>
    
```

① 共通参照レイヤーより粒度の細かい情報は、各DBの名前空間を使用

① 外部ヒートマップを参照

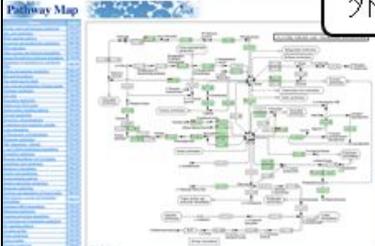
② 外部テーブルを参照

③ 外部パスウェイマップを参照

①

②

③



| No. | Probe ID | Gene Symbol | Title | P value | Fold change |
|-----|-------------|-------------|---|-----------|-------------|
| 1 | 209464_at | AURKB | Aurora kinase B | 0.0000002 | 3.59 |
| 2 | 202503_s_at | KIAA0101 | KIAA0101 | 0.0000007 | 5.24 |
| 3 | 202107_s_at | MCMC2 | minichromosome maintenance-deficient 2, mitotin | 0.0000012 | 4.33 |
| 4 | 218009_s_at | PRC1 | Protein regulator of cytokinesis 1 | 0.0000012 | 4.83 |

(図 G: OML 内のデータ構造 (2) Omics Annotation)

```

644 <ICD_annotation>
645 <personal_info>
646 <personal_description>
647 <ICOD.name>〇田〇男</ICOD.name>
648 <ICOD.nennrei>60代</ICOD.nennrei>
649 <ICOD.seibetsu>男</ICOD.seibetsu>
650 <ICOD.sinntyou>161.0</ICOD.sinntyou>
651 <ICOD.taijyuu>67.6</ICOD.taijyuu>
652 </personal_description>
653 <personal_dbref>
654 </personal_info>
655 <ICD_Classification_Annotation>
656 <ICD11_Code>2C90.1</ICD11_Code>
657 <ICD11_Concept_Title>Hepatocellular carcinoma of liver</ICD11_Concept_Title>
658 </ICD_Classification_Annotation>
659 <clinical_annotation>
660 <disease_disease_id_NCK="NCK-03125">
661 <ICOD.shubyoume> Hepatocellular carcinoma(肝細胞癌) </ICOD.shubyoume>
662 </disease>
663 <clinical_observation>
664 <subjective_findings>
665 <plan>
666 <objective_findings>
667 <medical_history>
668 <life_history>
669 </clinical_observation>
670 </clinical_annotation>
671 </ICD_annotation>
672 </OML>
    
```

患者基本情報
(各施設DB由来)

ICD11 への対応情報
(Transcription Module
でも使用)

患者診療情報
(各施設DB由来)

外部の疾患一般知識
(NCK.xml) 内の該当項目
へのポインタ

<Clinical observation> の中で
さらに詳細な
個々の診療情報を記述

次ページ

(図 H: OML 内のデータ構造 (3) ICD Annotation)

具体例

Clinical Observation >> Life History 記述例

```

834 <life_history>
835 <icod:sigoto_ryou>普通</icod:sigoto_ryou>
836 <icod:yuugai_bussitu_no_siyou>無</icod:yuugai_bussitu_no_siyou>
837 <icod:teiki_keikkou_sinnidann_no_jyusinn_jyokuyou>定期的</icod:teiki_
838 <icod:unnidou>しない</icod:unnidou>
839 <icod:unnidou_sports_no_keikenn>有</icod:unnidou_sports_no_keikenn>
840 <icod:suiminn_syougai>無</icod:suiminn_syougai>
841 <icod:kitsuenn>有</icod:kitsuenn>
842 <icod:kitsuenn_reki>10</icod:kitsuenn_reki>
843 <icod:honsuu>15.0</icod:honsuu>
844 <icod:kitsuenn_no_jikann>常時</icod:kitsuenn_no_jikann>
845 <icod:kinnenn_reki>0.0</icod:kinnenn_reki>
846 <icod:innsyu>有</icod:innsyu>
847 <icod:innsyu_reki>15</icod:innsyu_reki>
848 <icod:hinndo>3</icod:hinndo>
849 <icod:alcohol_no_syuru>ビール</icod:alcohol_no_syuru>
850 <icod:innsyu_ryou>5</icod:innsyu_ryou>
851 <icod:nomu_kikai_no_ooi_nomimono>日本茶、野菜ジュース</icod:nomu_kikai_no_ooi_nomimono>
852 <icod:kennkou_syokuhinn_supplement>無</icod:kennkou_syokuhinn_supplement>
853 <icod:syokuji_kaisuu>3.0</icod:syokuji_kaisuu>
854 <icod:syokuji_jikann>規則的</icod:syokuji_jikann>
855 <icod:syokuji_sessyu_jikann>普通</icod:syokuji_sessyu_jikann>
856 <icod:syokuji_no_sessyu_houhou_ryou>普通</icod:syokuji_no_sessyu_houhou_ryou>
857 <icod:hennsyoku_noumu>無</icod:hennsyoku_noumu>
858 <icod:tokuni_sukina_syokuhinn_ryouri_mei>寿司</icod:tokuni_sukina_syokuhinn_ryouri_mei>
859 <icod:nikurui_sessyu_ryou>普通</icod:nikurui_sessyu_ryou>
860 <icod:zyourui_sessyu_ryou>普通</icod:zyourui_sessyu_ryou>
861 <icod:yasai_sessyu_ryou>普通</icod:yasai_sessyu_ryou>
862 <icod:kudamono_sessyu_ryou>普通</icod:kudamono_sessyu_ryou>
863 <icod:ajituke_ya_konomi>薄</icod:ajituke_ya_konomi>
864 <icod:karai_mono_sisekibutu_no_sessyu>好まない</icod:karai_mono_sisekibutu_no_sessyu>
865 <icod:kannsyoku>無</icod:kannsyoku>
866 <icod:sikouno_hennka>無</icod:sikouno_hennka>
867 <icod:sikou_no_jiki>無</icod:sikou_no_jiki>
868 <icod:haibenn_syuuikann>有</icod:haibenn_syuuikann>
869 <icod:tuubenn_jiyou>無</icod:tuubenn_jiyou>
870 <icod:geketo_no_umu>有</icod:geketo_no_umu>
871 <icod:geri_no_umu>無</icod:geri_no_umu>
872 <icod:bennpi_no_umu>無</icod:bennpi_no_umu>
873 <icod:gezaishiyou_no_umu>無</icod:gezaishiyou_no_umu>
874 <icod:ji_no_umu>有</icod:ji_no_umu>
875 <icod:gis>下前歯2本以外</icod:gis> |
876 <icod:gis_i_ni_tuite>入れ歯</icod:gis_i_ni_tuite>
877 <icod:kounaienn_no_keikenn>有</icod:kounaienn_no_keikenn>
878 <icod:kounaienn_ninaru_hinndo>少ない</icod:kounaienn_ninaru_hinndo>
879 <icod:sisou_nourou>有</icod:sisou_nourou>
880 <icod:siseki_no_jyokyo_keikenn>無</icod:siseki_no_jyokyo_keikenn>
881 <icod:siseki_no_jyokyo_hinndo>少ない</icod:siseki_no_jyokyo_hinndo>
882 <icod:hamigaki_no_kaisuu>2</icod:hamigaki_no_kaisuu>
883 <icod:gum_wash_nado_no_siyou>無</icod:gum_wash_nado_no_siyou>
884 </life_history>

```

共通参照レイヤーより粒度の細かい情報は、各DBの名前空間を使用

<icod:運動>しない

<icod:喫煙歴>10年

<icod:肉類摂取量>普通

<icod:便秘の有無>無し

(図 I: Clinical Observation の下位項目 Life History 記述例)

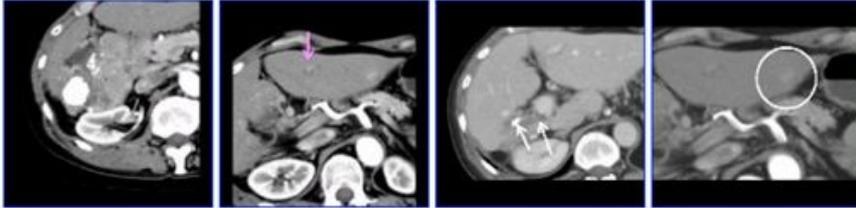
具体例

Clinical observation >> 画像検査情報 記述例

757
758
759
760
761
762
763
764
765
766
767

```
<diagnostic_image>
  <icod:angiography_kiensa_bui/>
  <icod:angiography_zouei_zai/>
  <icod:angiography_syokenn/>
  <icod:angiography_comment/>
  <icod:ct_record_id:figure_ct00001/</icod:ct_record_id>
  <icod:ct_kiensa_bui>肝・胆・膵・脾</icod:ct_kiensa_bui>
  <icod:ct_zouei_zai/</icod:ct_zouei_zai>
  <icod:ct_syokenn/</icod:ct_syokenn>
  <icod:ct_comment>HCC:萎縮した右葉から大きく下部背側に突出。S3にdaughter lesion。門脈右枝は起始部で閉塞するが、血栓による閉塞か、腫瘍栓か判断困難</icod:ct_comment>
</diagnostic_image>
```

(例) CT所見の記述と、外部画像ファイルの参照ポインタ



768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784

```
<endoscopic_screening>
  <icod:naisikyou_youhou_record_id:figure_n000001/</icod:naisikyou_youhou_record_id>
  <icod:naisikyou_youhou_sindann>胃隆起性病変(0-Ba疑い)、萎縮性胃炎</icod:naisikyou_youhou_sindann>
  <icod:naisikyou_youhou_comment>なし</icod:naisikyou_youhou_comment>
  <icod:naisikyou_youhou_bui>咽喉</icod:naisikyou_youhou_bui>
  <icod:naisikyou_youhou_bui_syokenn/>正常</icod:naisikyou_youhou_bui_syokenn>
  <icod:naisikyou_youhou_bui>食道</icod:naisikyou_youhou_bui>
  <icod:naisikyou_youhou_bui_syokenn/>正常、静脈瘤なし</icod:naisikyou_youhou_bui_syokenn>
  <icod:naisikyou_youhou_bui>胃</icod:naisikyou_youhou_bui>
  <icod:naisikyou_youhou_bui_syokenn/>
  <icod:naisikyou_youhou_bui>十二指腸</icod:naisikyou_youhou_bui>
  <icod:naisikyou_youhou_bui_syokenn/>正常</icod:naisikyou_youhou_bui_syokenn>
</endoscopic_screening>
```

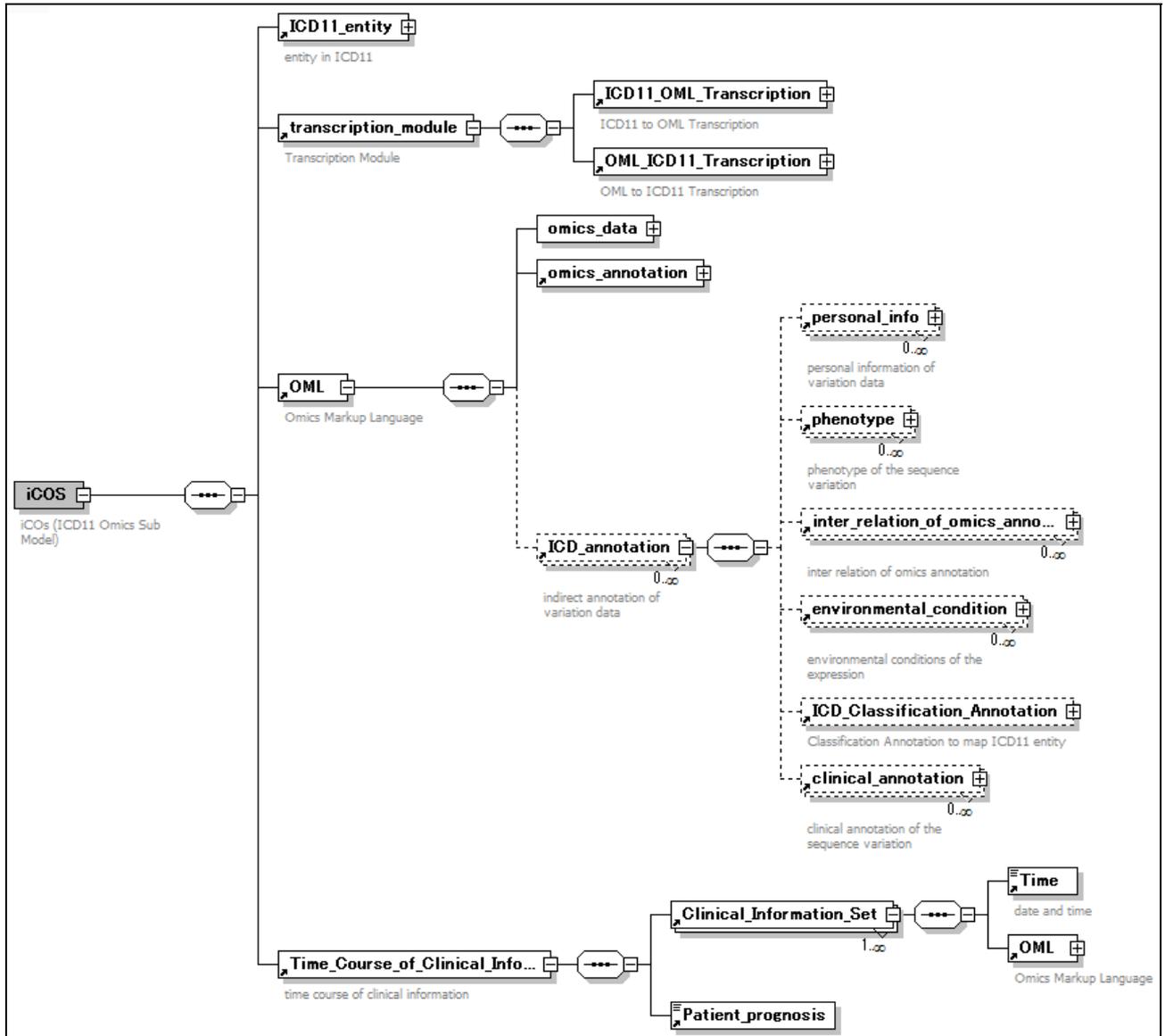
痔瘻:静脈瘤なし。高悪性腫瘍性結腸憩室あり。後上部小腸(後壁部)に発赤を伴う径10mm程度の隆起性病変あり。
 胃小凹は乱れ、精確は明確で中央はやや陥凹している。0-Baを疑う。隆起から生検1・2。進行。
 前胃結膜は萎縮と軽度の腸上皮化生を伴う。その小腸(胃)に小凹(aneurodysplasiaあり)。前庭部は発赤あり。

(例) 内視鏡検査結果の記述と、外部画像ファイルの参照ポインタ

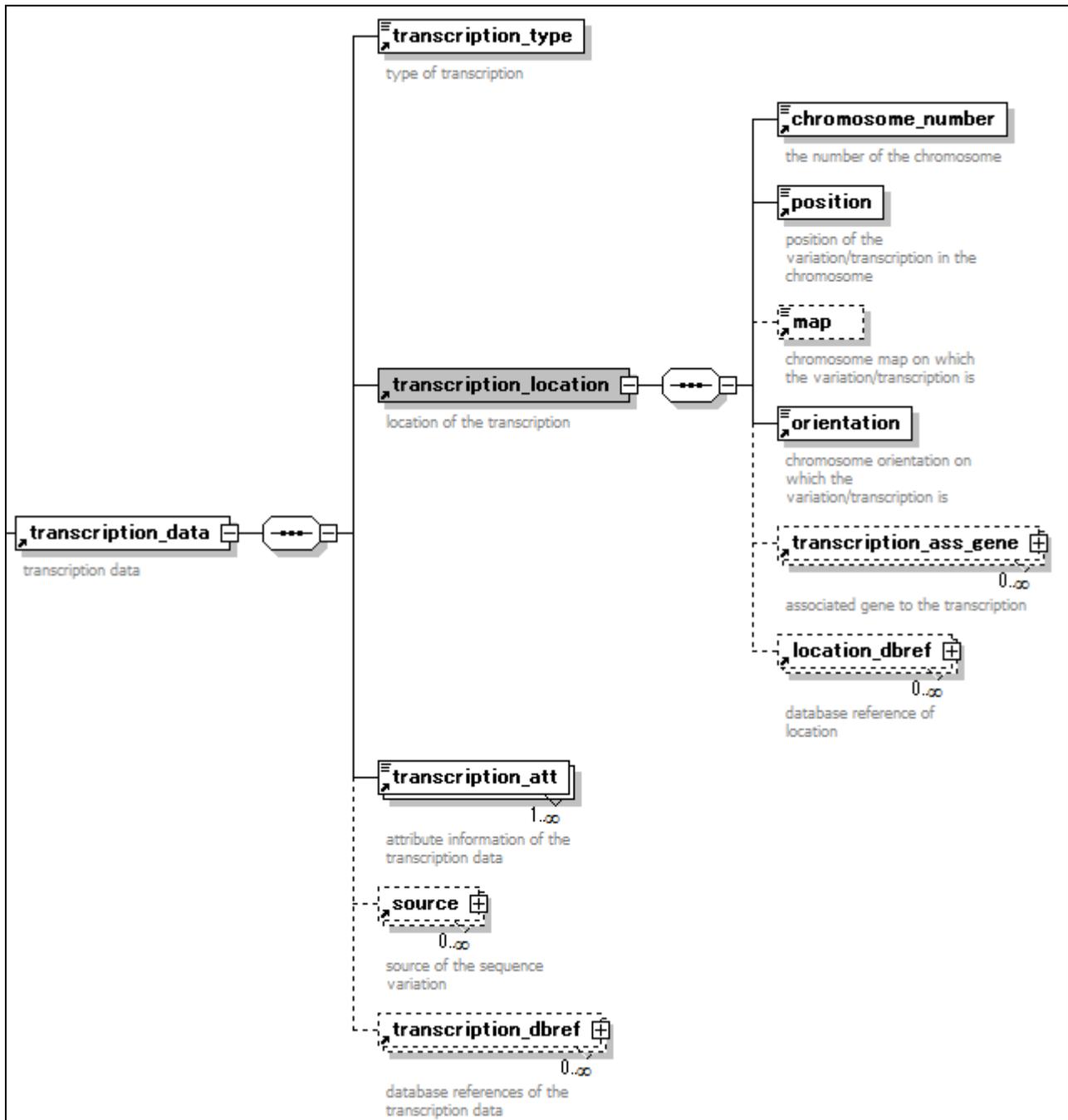


(図 J: Clinical Observation 中の下位項目 画像検査情報記述例)

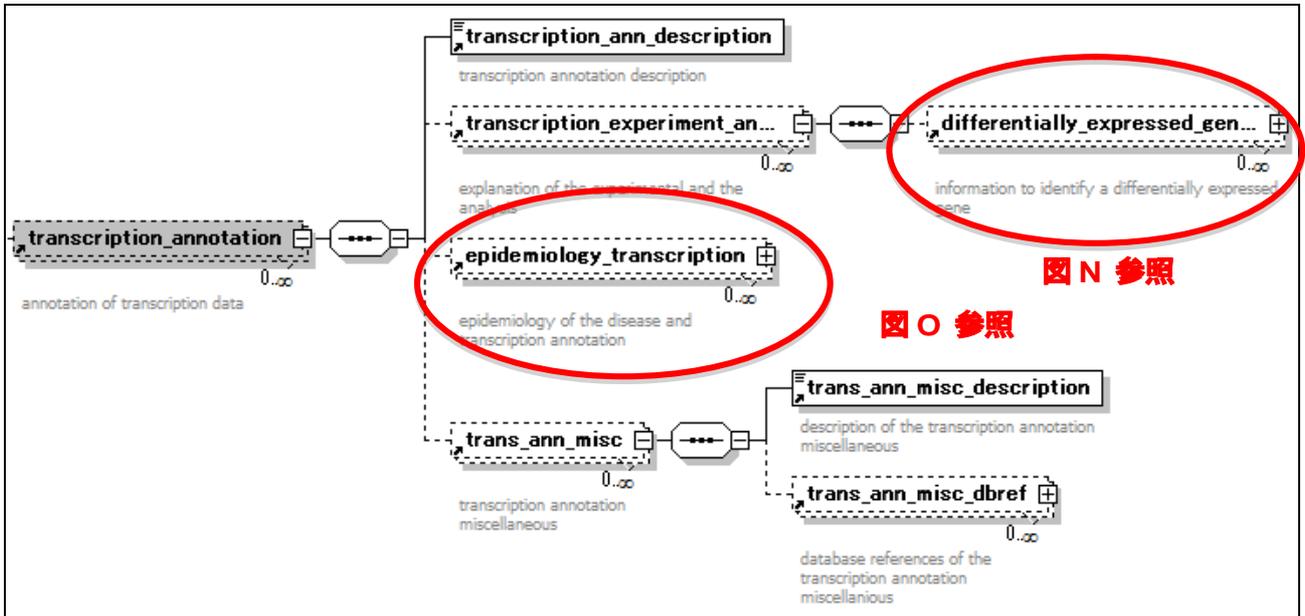
(以下は、昨年度までに構築した iCOS モデルの参考資料です)



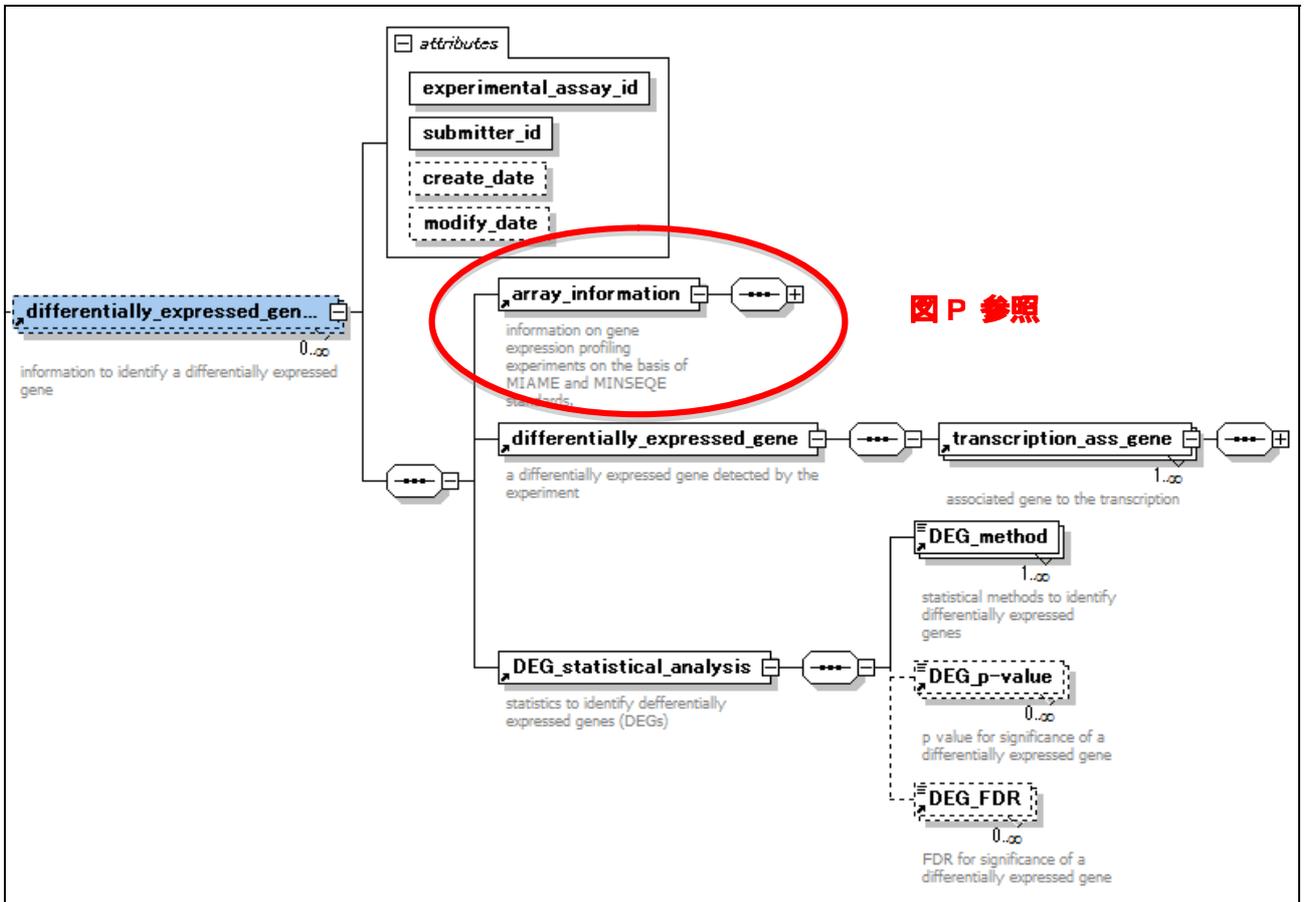
(図 K. iCOS モデルの上位構造と “Time course of clinical information” の位置づけ)



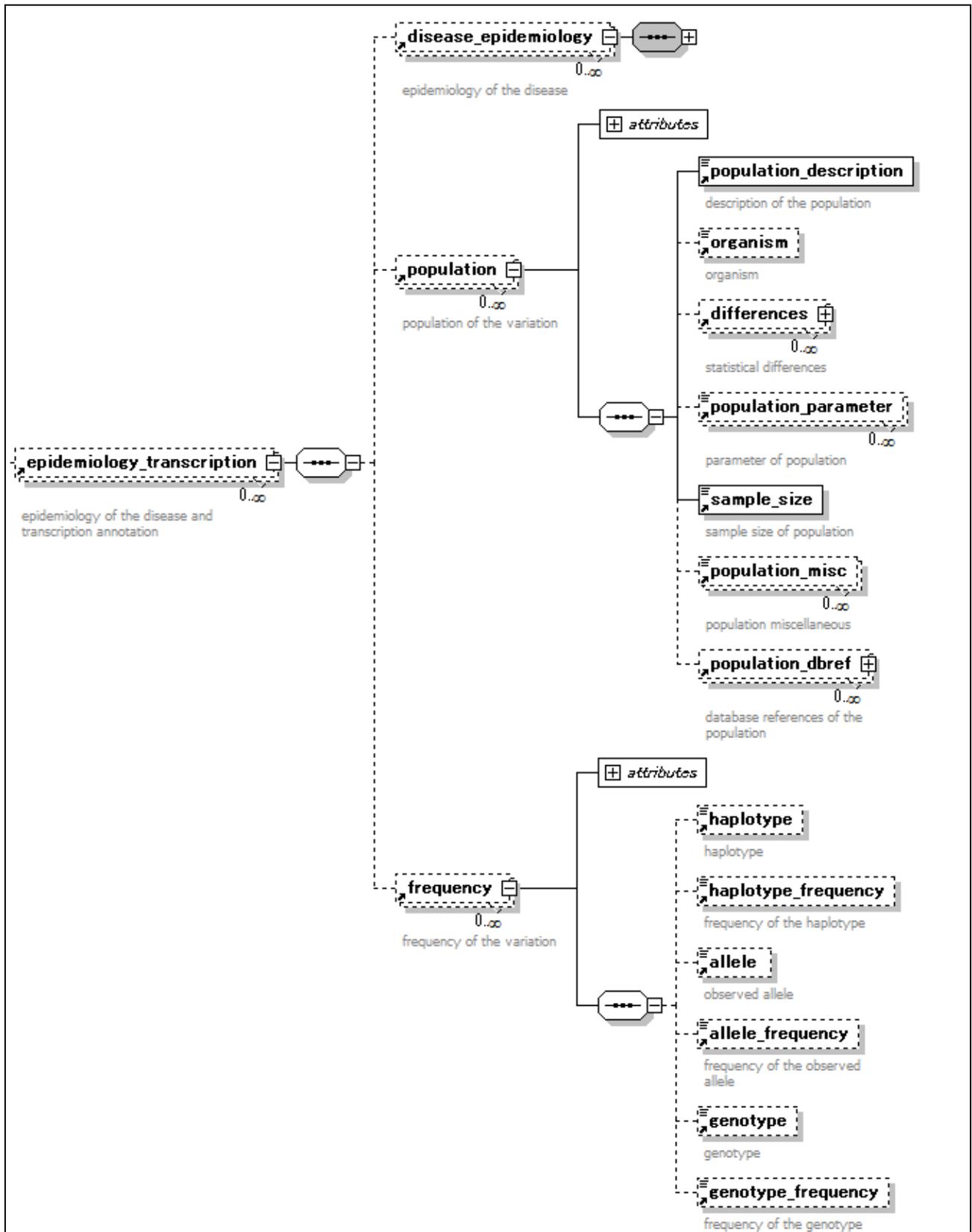
(図 L. Transcription_Data の詳細構造 (2 階層まで展開))



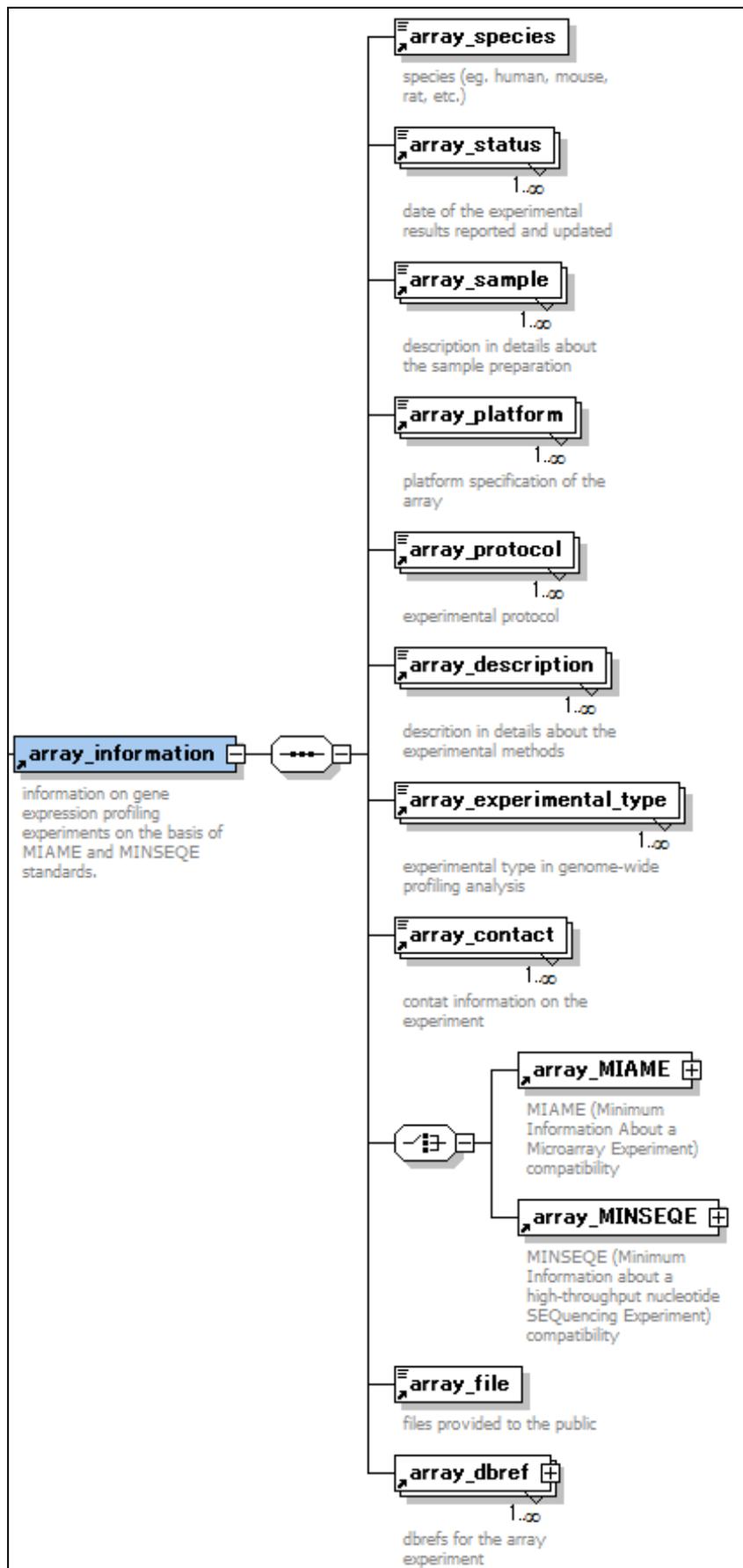
(図 M. Transcription_Annotation の詳細構造 (2 階層まで展開))



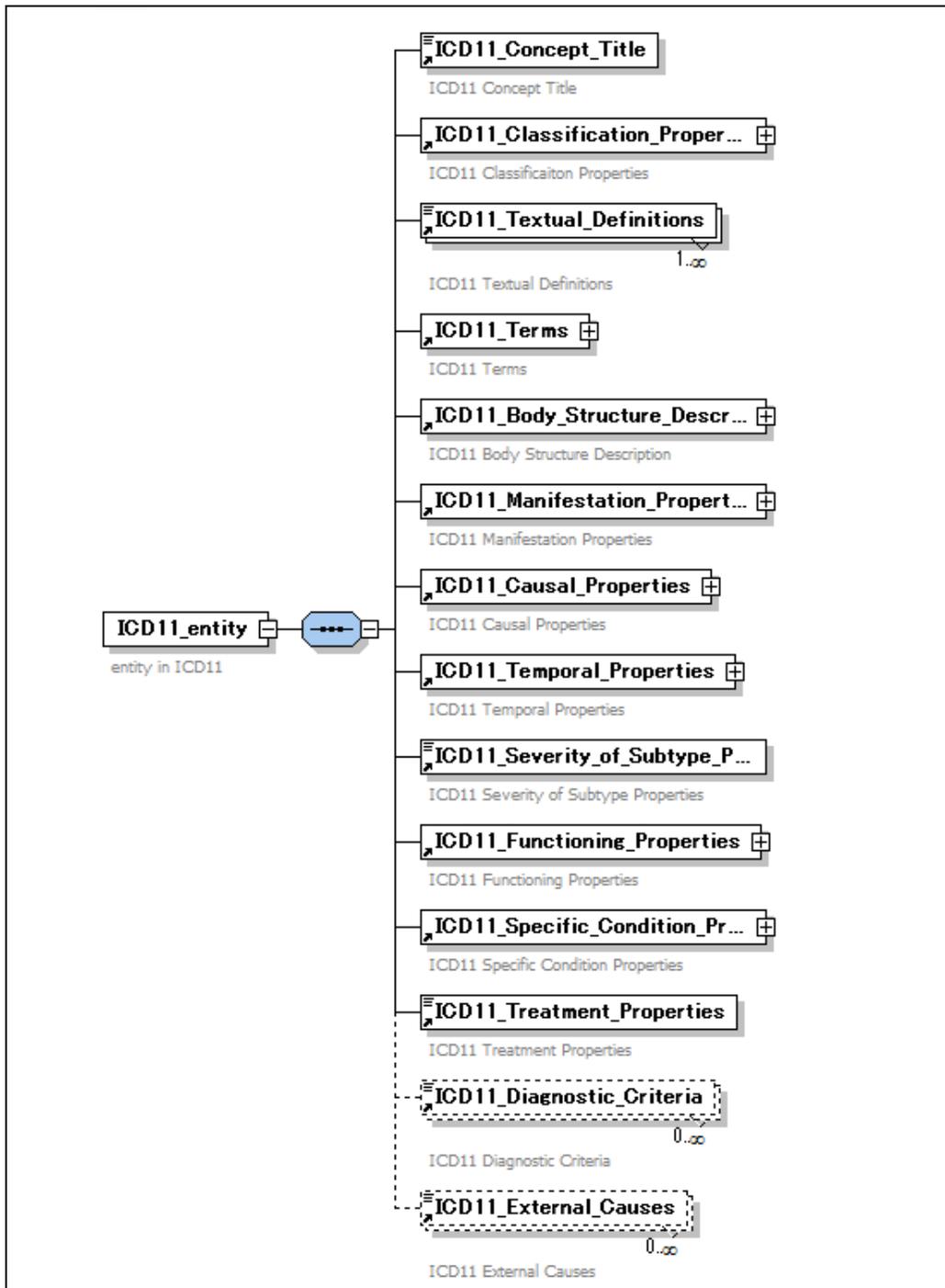
(図 N. “differentially_expressed_gene_analysis” の詳細構造 (2 階層まで展開))



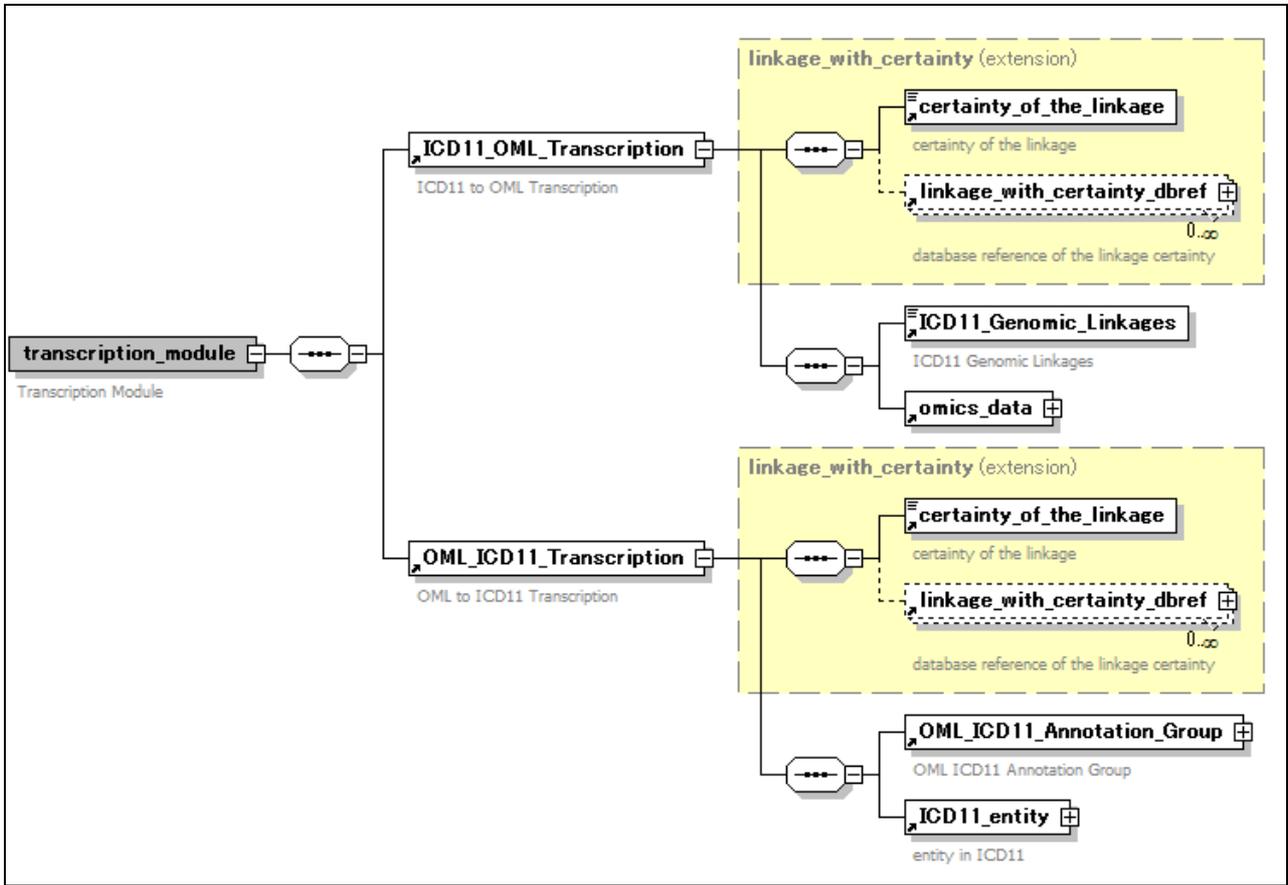
(図 O “epidemiology_transcription” の詳細構造 (2 階層まで展開))



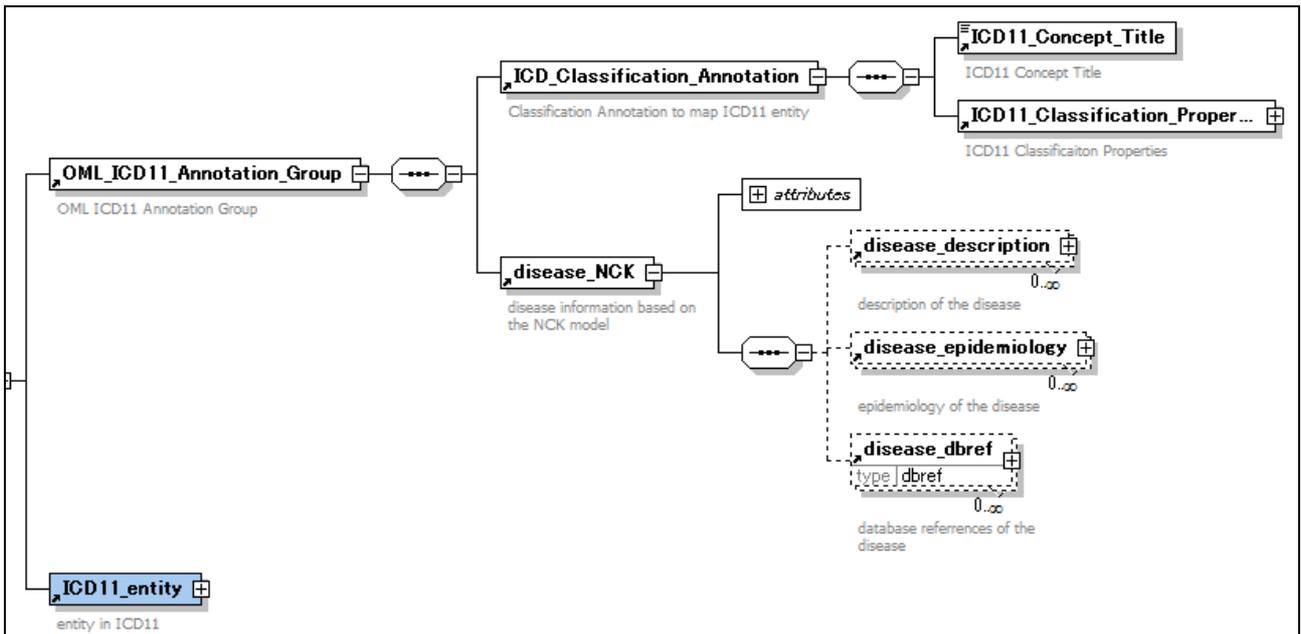
(図 P. “array_information” の詳細構造)



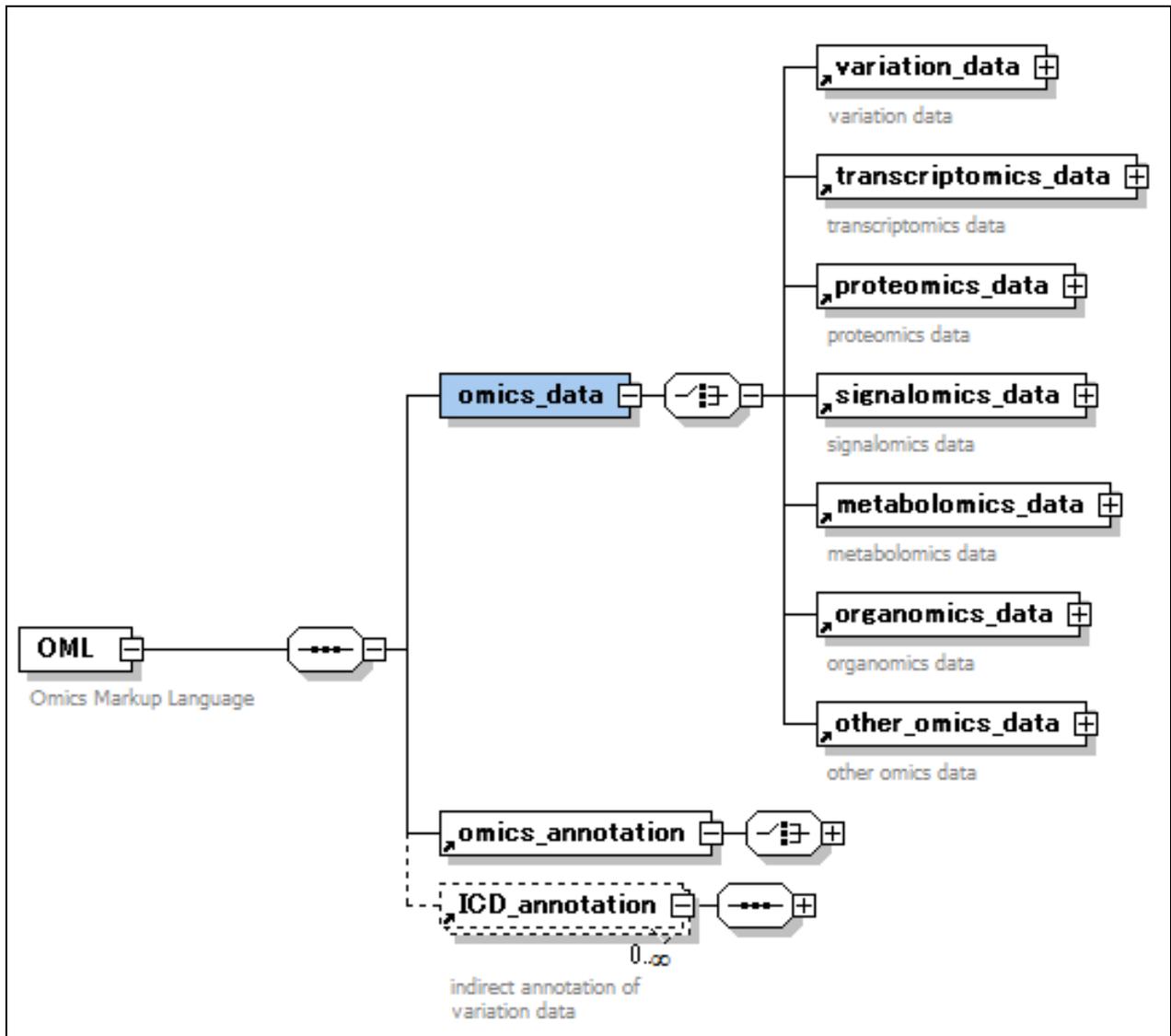
(図 Q . ICD11_entity の概要 (第 1 階層まで展開))



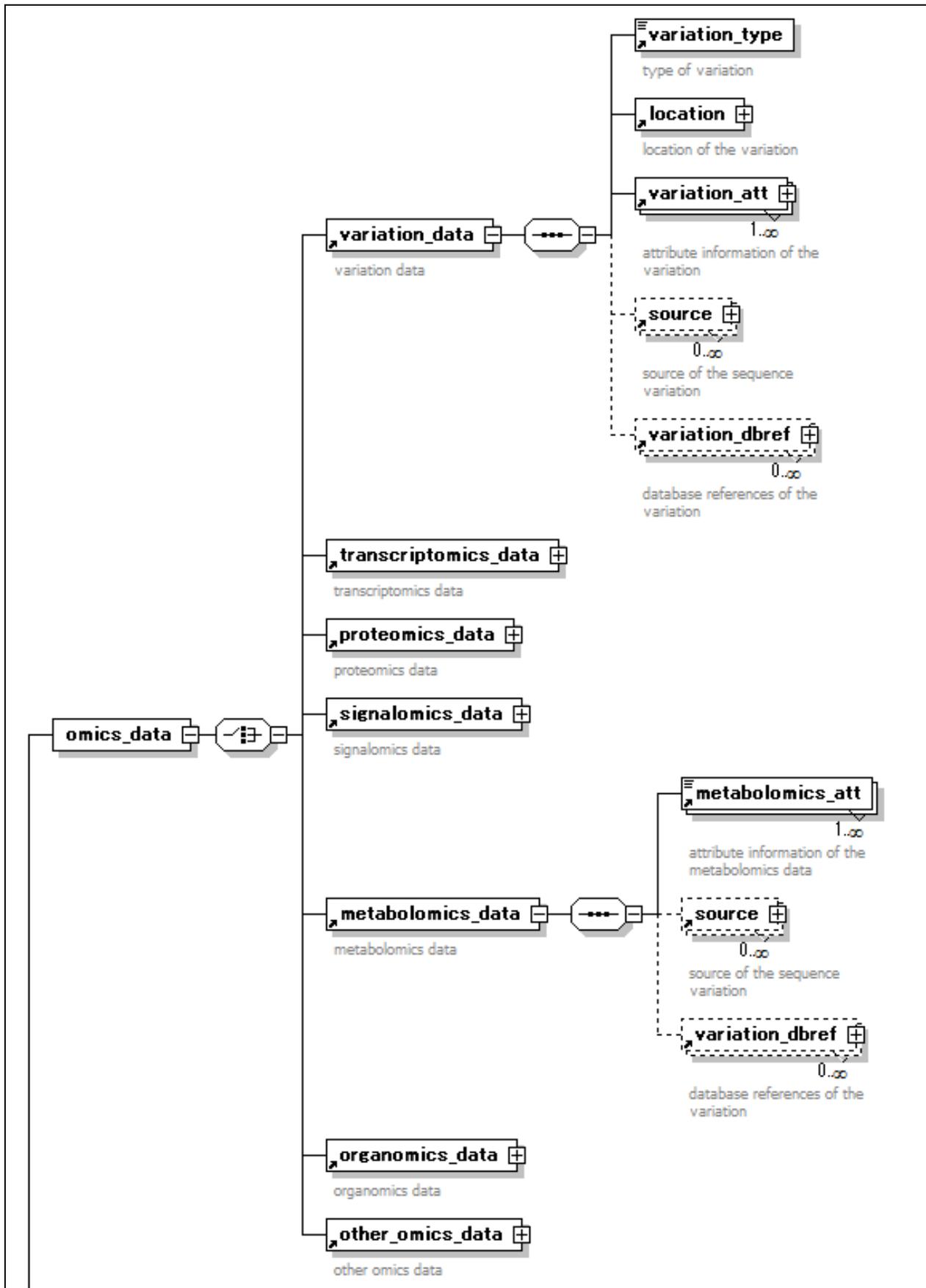
(図 R . Transcription_module (ICD11 ⇔ OML 双方向転写モジュール) の概要)
 (2 階層目まで展開)



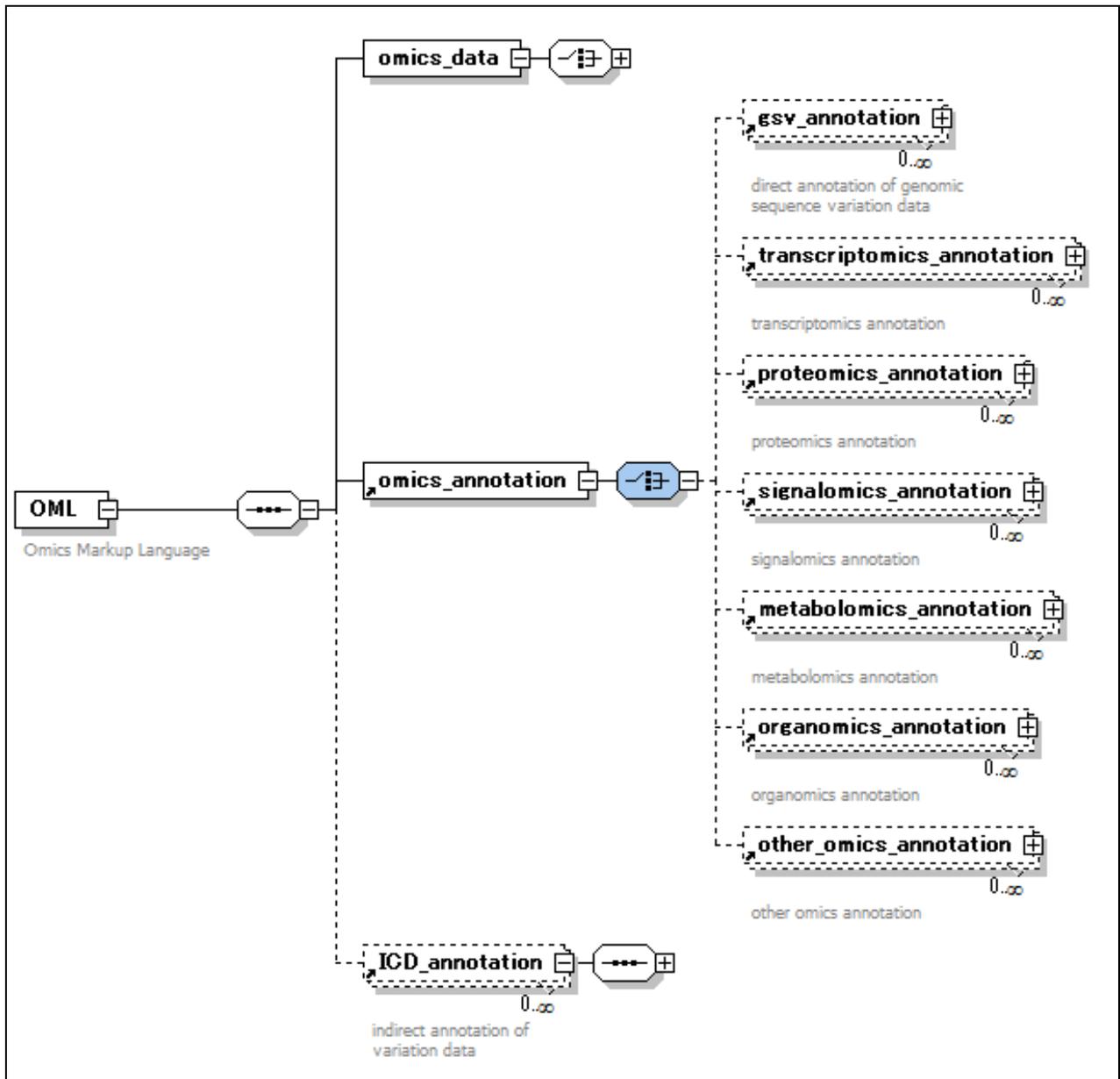
(図 S . OML_ICD11_Annotation_Group の構造)



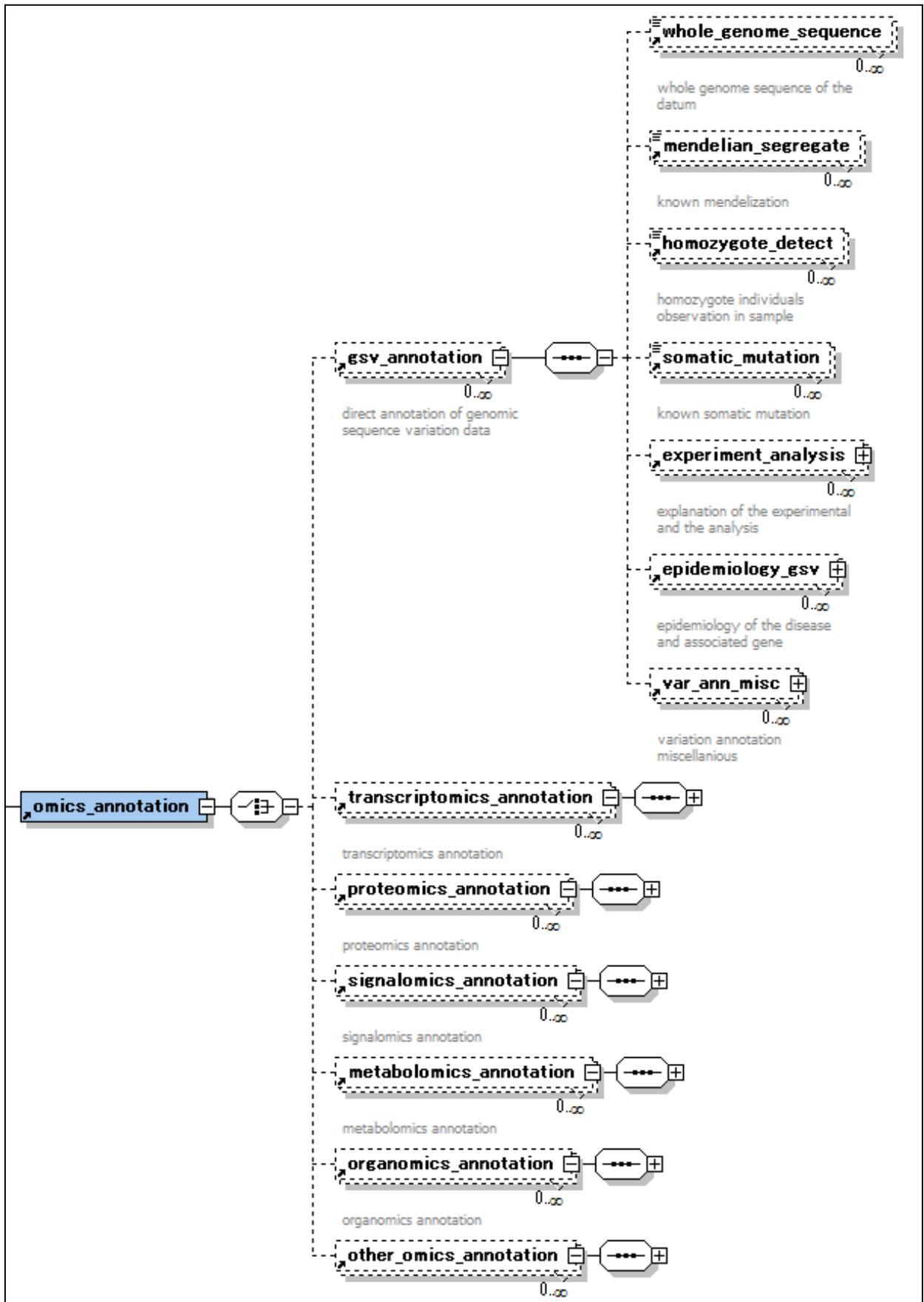
(図 T . OML (Omics Markup Language) と Omics_data の概要)



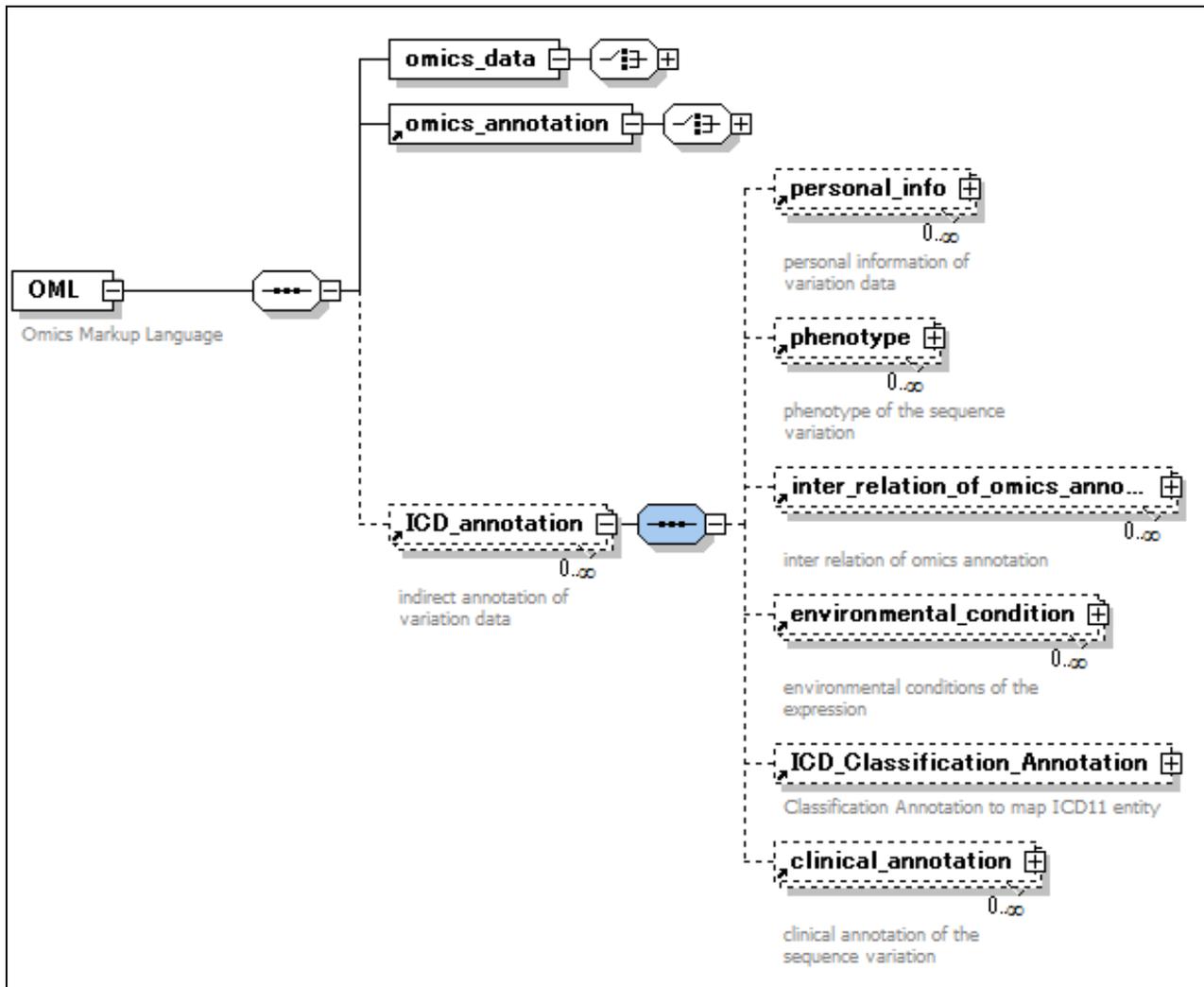
(図 U . OML > Omics_data 以下の構造)



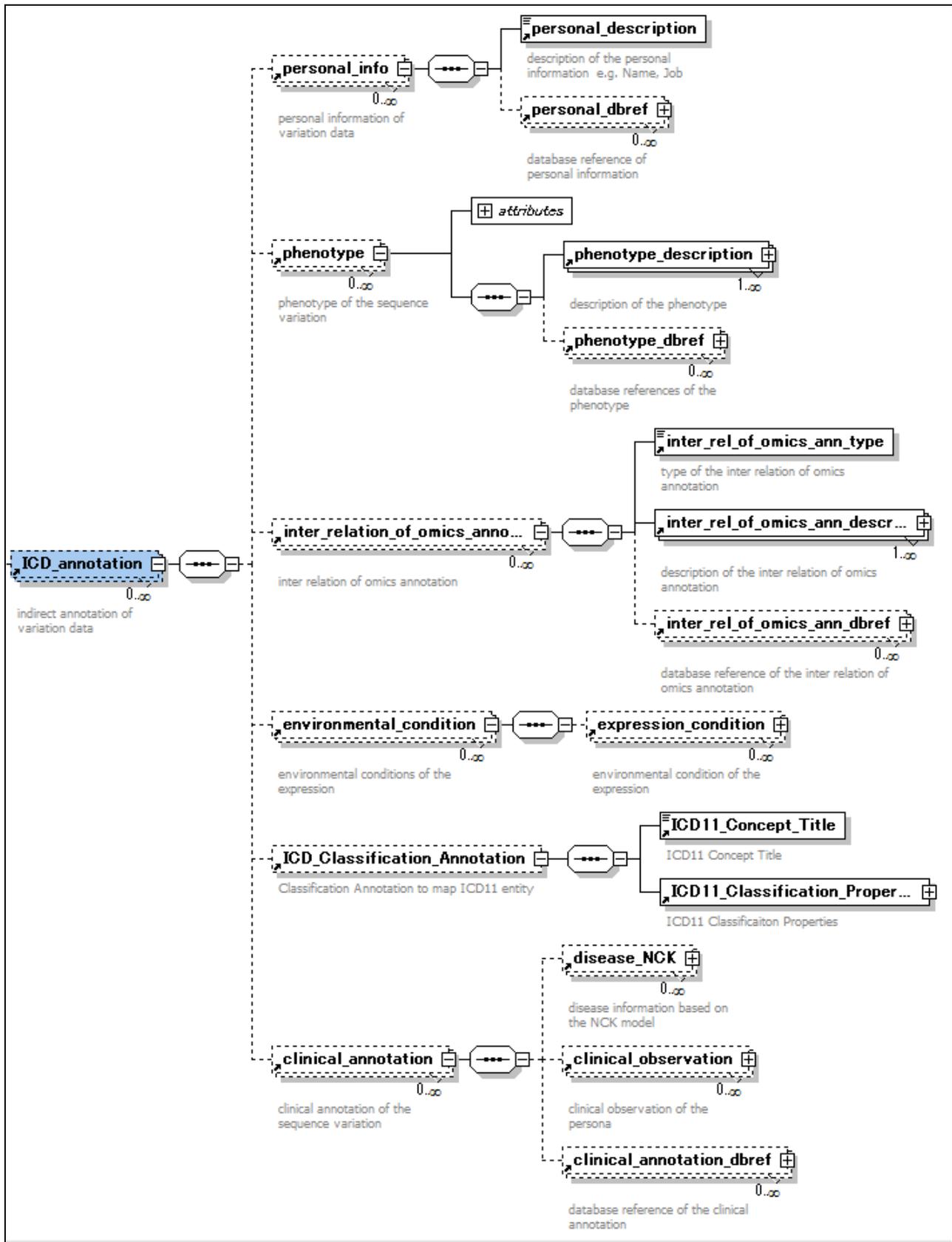
(図 V . OML (Omics Markup Language) と Omics_annotation の概要)



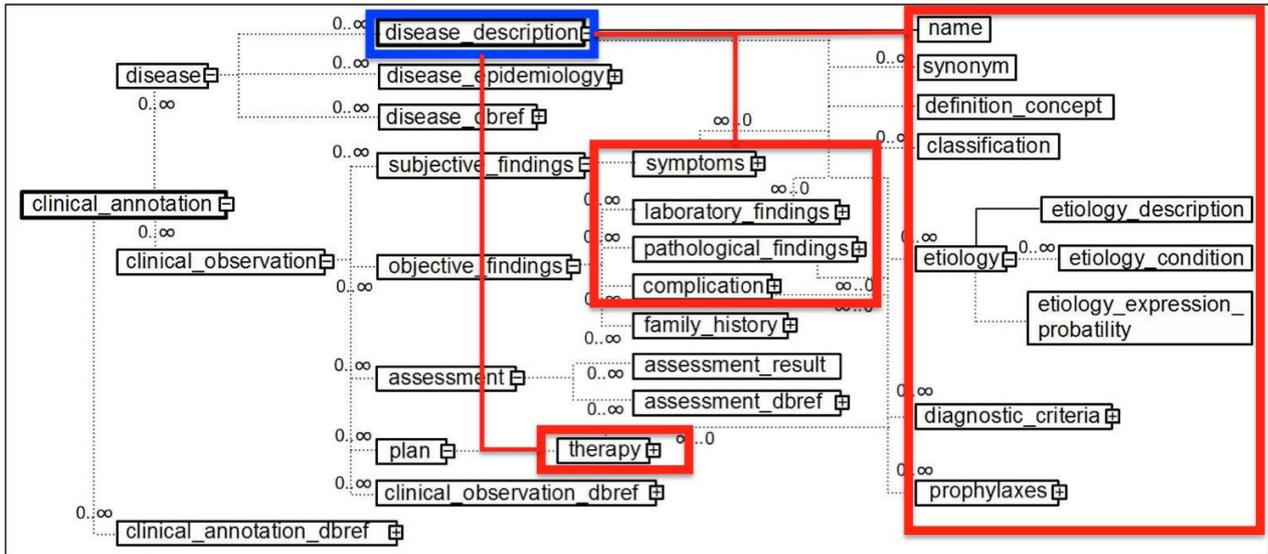
(図 W . OML > Omics_annotation 以下の構造)



(図 X . OML (Omics Markup Language) と ICD_annotation の概要)



(図 Y . OML > ICD_annotation 以下の詳細)



(図 Z . disease_NCK (GSVML) 内の詳細)