

201302002B

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

「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」

平成 24 年度～25 年度 総合研究報告書

研究代表者 中谷 純

平成 26（2014年）5月

目 次

I. 総合研究報告		
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	-----	1
中谷純		
II. 研究成果の刊行に関する一覧表	-----	35

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）
総合研究報告書

「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 本研究では、2015年に改訂が予定されている WHO-ICD11 のコンテンツモデルに組み込むことのできる臨床 オミックス情報モデル(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。iCOS は、本研究に先立ち、平成 22 年度厚生労働科学研究費補助金(統計情報総合研究事業)「ICD11 におけるオミックス情報モデルの研究」において作成された情報サブモデルであり、ISO において正式に国際標準(IS)と認められた日本発進の技術である GSVML を応用し ICD11 コンテンツモデルに組み込むことのできる臨床オミックス情報モデルとして開発された日本発進の技術である。平成 24 年度は、臨床オミックス情報モデルが使われる場合の利用形態を情報学的観点から整理して、ユースケースの作成を行う。次に、網羅的かつ簡易的なデータを用いて多角的な観点から実証検討を進め、WHO ICD11 TAG-HIM との情報交換などを行いながら、iCOS の網羅的概形実証検討を行う。平成 25 年度は、全ゲノムデータ、階層オミックスデータ、臨床詳細情報などを入力データとして使用し、iCOS の限界性能を明らかにする。また、iCOS を ICD11 の中で利用するための利用指針の作成を行いつつ、必要に応じて iCOS の改良を行う。事実上の国際標準である ICD11 において追加されるオミックス情報カテゴリーをどのように記述するかは、今後の国際医学研究を左右する大きな課題である。日本が世界に先駆け国際標準として樹立した ISO 国際標準 GSVML 情報モデルを応用する形で提示した ICD11 オミックス情報モデル iCOS は、日本発進の新技術として誇ることのできる大きな特色であり、独創的な点である。本研究はオミックス医学という分子生物医学の集大成ともいえる医学分野を現在の世界保健の主流である疾病分類に組み込むために必須のものであり、ICD11 に整合性の取れたオミックス情報モデル iCOS を現実的な視点で利用するための実証検討を行うことは、ICD 改訂に資する、大きな意義のある時宜にかなった現実的な研究である。

研究分担者

田中 博 東京医科歯科大学 教授
今井 健 東京大学 助教

A. 研究目的

ICD11 改訂に資するため、平成 22 年度厚生労働科学研究費補助金(統計情報総合研究事業)「ICD11 におけるオミックス情報モデルの研究」の成果である ICD11 臨床オミックス情報モデル(ICD11 Omics Sub information model: iCOS)の妥当性に関する多角的実証研究を行う。

B. 研究方法

本研究に先立って既に構築済みである平成 22 年度厚生労働科学研究費補助金(統計情報総合研究事業)「ICD11 におけるオミックス情報モデルの研究」の成果である ICD11 オミックスサブ情報モデル(ICD11 Omics Subinformation model: iCOs)を検討材料として用いた。iCOs モデルには WHO-ICD11 の特性記述モデル(Content Model)が含まれているが、その最新動向については厚生労働省 ICD11 国内内科 TAG 検討会に参加し、情報収集を行った。

次に iCOs モデルの妥当性検証を行うにあたり必要不可欠なユースケースを設定した。さらに iCOs モデルによる「実臨床オミックスデータベースの記述可能性」を検証するため、東京

医科歯科大学の網羅的分子病態データベース(iCOD)を例として用い、iCOs モデルとの情報項目マッピングを行った。その際には、先に設定したユースケースに基づき、iCOs で網羅すべき汎用的な情報粒度と、各個別データベースに固有とするべき情報粒度との線引き(共通参照レイヤー)を設定した。これは他の既存リソース(GO 等)も参考にしながら、オミックス情報・臨床情報の全ての情報項目について、分野専門家の協力の下に行った。

得られたマッピング結果から、iCOs モデルの不備を明らかにし、モデルの拡充を行った。修正した iCOS モデルは Altova XML Spy を用いて XML Schema として記述した。

拡充を行った 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(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)が含まれ、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、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

1 ユースケースの設定

iCOs モデルがどうあるべきかを考える上でユースケースの設定は必要不可欠である。本年度は以下のようなユースケース概要を設定し、全体設計を行っていくこととした。

- 患者のオミックス情報と診療情報の両方を蓄積した「臨床オミックス研究データベース」を多施設間で統合的に解析する際に「共通参照情報モデル」として用いる。
- iCOs モデルの XML Schema に基づいて、実際に上記データベースを格納する。
- 2015 年の WHO-ICD11 リリースを見据え、将来的に ICD11 で記述される診療情報・疾患知識を上記データベースに格納することができる。
- また ICD11 と連携して利用するための利用指針を提供する。

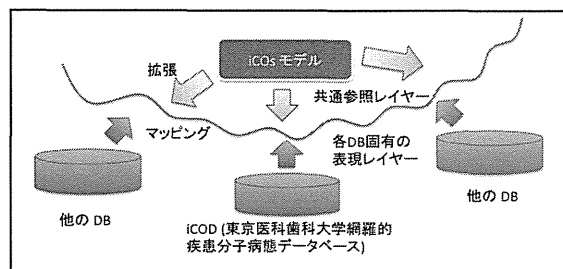
2 “Mapping principle” の設定

1 のようなユースケースを考えた場合、iCOs モデルは多施設の臨床オミックスデータベースを対象に横断的に解析するための「共通参照モデル」として機能することが求められる。従って特定のデータベースのスキーマに高度に依

存しすぎるとは共通性が失われる問題が生じる。そこで、本研究では実臨床データベースとの情報項目のマッピングを行う際の原則として、共通とすべき情報粒度の境界を定める“2-layer モデル”を採用することとした(図 A)。

これは CEN/ISO13606 あるいは OpenEHR 規格における多施設間の診療情報連携においても採用されている考え方であり、具体的な方針は以下の通りである。

- 汎用的な情報エンティティと、各データベース固有の情報エンティティの情報粒度を区分(分野の専門家で議論して決定)
- 必要があれば、汎用的なレベル(共通参照レイヤー)の情報粒度までは iCOs を拡張
- 各DB固有のエンティティは iCOs ではカバーしないものとし、各DBからの(データ出力時の)マッピングに任せる



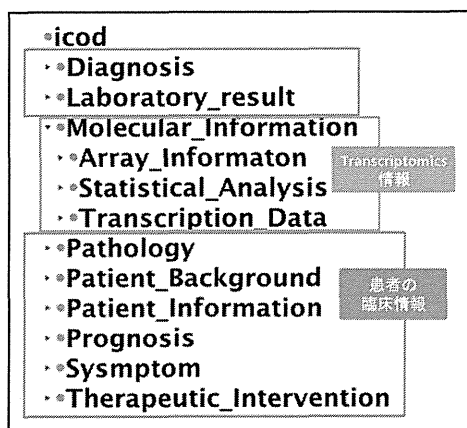
(図 A) 2-layer モデルの考え方

3 実データリソースとのマッピングに基づく記述可能性検討

iCOs モデルによる「実臨床オミックスデータベースの記述可能性」を検証するため、東京医科歯科大学の網羅的分子病態データベース(iCOD)を例として用い、「Mapping principle」である 2-layer モデルに則り、iCOs との情報項目マッピングを行った。以下、最初に両者の構造概要を簡単に示し、次に情報項目のカテゴリ毎にマッピング対応の詳細について示す。

3.1 iCOD データベース構造

今回、iCOs とマッピングする実データリソースとして使用した東京医科歯科大学の網羅的分子病態データベース(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。まず、このテーブルの各項目を全て抽出し、検査、症状といった意味的なカテゴリに従って、階層分類を行った。最上位は図 2 に示すように Transcriptomics 分野のオミックス情報である “Molecular_information”



(図 B) iCOD データベース情報項目 (上位)

と、その検体の患者の臨床情報である “Diagnosis”, “Laboratory_result”, “Pathology”, “Symptom” といった内容から構成されており、最下層まで展開すると全部で 24 個の情報項目が存在していた。これらのデータベース情報を iCOs モデルで記述することは、iCOD 中の特定のデータベース項目の値を、その項目と対応する iCOs の情報エンティティに格納するという他にない。従って、以降はこれらの情報項目と iCOs モデルの情報エンティティとの対応関係を調査した。

3.2 iCOs モデル構造

iCOs は日本発の国際規格である ISO 25720 Genomic Sequence Variation Markup Language (GSVML) 情報モデルを応用して作られた、臨床オミックス情報の記述モデルであり、オミックス情報、検体の患者の臨床情報が ICD11 のコンテンツモデルと連携した形で記述できるように設計されている。概要を(別添)図 1 に示す。

iCOs で記述される情報は(別添)図 2 のようにオミックス情報、患者の臨床情報、外部知識の 3 つに大きく区分けすることができるが、東京医科歯科大学の iCOD データベース項目と対応するのは “Omics data”, “Omics annotation”, “ICD annotation” の各下位エンティティとなっていた。

iCOs モデル全体の詳細は、(別添)図 3~18 を参照されたい。

3.3 iCOD データ項目のマッピング

iCOD データベース項目は、Primary information, Symptom, Therapeutic intervention, Laboratory result 等のカテゴリの下位に多数の項目が存在していたが、2Layer モデルに従い「多施設の臨床オミックスデータベースに共通とするべき項目」(共通参照レイヤー)を分野専門家と協議し選定した。結果は以下の通りである。

- (1) [Diagnosis カテゴリ]… “diagnosis”
- (2) [Primary information カテゴリ]… “Age”, “Sex”

- (3) [Symptom カテゴリ]… “Symptom”
- (4) [Therapeutic intervention カテゴリ]… “Drug”, “Radiographic intervention”, “Surgical intervention”
- (5) [Laboratory result カテゴリ]… “Blood test”, “Diagnostic imaging”, “Endoscopic screening”, “Physique”
- (6) [Molecular information カテゴリ]… “Statistical methods”, “Experimental methods”, “kinds of chips used”, “sampling”
- (7) [Pathology カテゴリ]… “Macroscopic finding”, “Tissue finding”
- (8) [Patient background カテゴリ]… “Medical history”, “Occupation diet and lifecycle”
- (9) [Prognosis カテゴリ]… “prognosis”

これ以上詳細な情報粒度は各データベース固有項目とし、iCOs に格納するには、例えば <blood test>

```
<TP> xxx </TP>
<ALB> yyy </ALB>
```

</blood test>

のような XML 形式に構造化し、“blood test” エンティティ内に記述することを想定している。

次に、これらの共通参照レイヤー項目を iCOs のエンティティとマッピングした結果を(別添)表 1 に示す。例えば iCOD の “diagnosis” であれば、iCOs 上の “disease NCK” という項目に対応している。一方「拡張項目」として色づけされているものは、既存の iCOs の情報粒度が足りず、共通参照レイヤーの情報粒度に対応するために iCOs 側に追加することで、結果的にマッピングが可能となったエンティティを示している。例えば従来 iCOs には “laboratory findings” の粒度までしか項目が存在していなかったが、今回 “blood test”, “diagnostic imaging” 等を追加することでマッピングを行っている。結果、既存の iCOs モデルの考え方を大きく変更することなく、少数の項目を追加するだけで iCOD データベース項目(共通参照レイヤーの情報粒度)とのマッピングが可能であることが判明した。

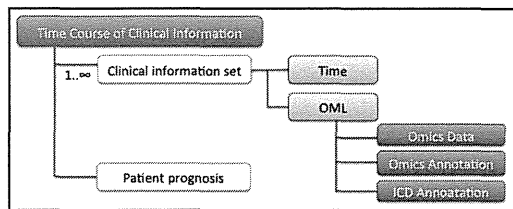
4 iCOs に追加した主なモジュール

既存の iCOs モデルに追加したモジュール(エンティティ群)の一覧は(別添)表 A に示した通りで、所見の細分化など軽微な変更がほとんどであったが、以下ではその中でも特に主要な 2 つの追加モジュールについて述べる。

4.1 Time course of clinical information

iCOs は元々、静的な情報の記述を想定しており、診療録の時系列記録は含んでいなかった。しかし、臨床オミックス研究においては時系列

変化情報も重要になる場合があることから、臨床情報 (ICD_annotation), オミックス情報 (Omics_data, Omics_annotation) の時系列変化を記述するためのモジュールを追加した。その構造を図 C に示す。時系列情報のみならず、予後/最終的なアウトカム等も記述できるように設計されている。XML Schema では (別添) 図 C に示す通り、トップノード "iCOs" の直下に位置づけられている。



(図 C. Time course of clinical information)

4.2 Transcription data / annotation

今回マッピングに用いた実臨床オミックスデータベース iCOD はオミックスの中でも特に Transcriptomics 領域の情報を対象としている。一方 iCOs は GSVML (Genomic sequence variation markup language) モデルから発展してきているため、Transcriptomics 領域の情報項目があまり含まれていなかった。そこで今回、Gene Ontology や SNOMED-CT など既存の医学・分子生物学知識リソースにおける記述モデルの考え方も参照しながら、"Transcription data" 並びに "Transcription annotation" の下位構造を (別添) 図 4~8 に示すように記述した。

"Transcription data" (別添 図 4) は "Omics data" の下位に位置し、transcription に関する Raw Data を記述するためのモジュールであり、"transcription type", "transcription location" などの情報を格納するように設計されている。

一方、"Transcription annotation" (別添 図 5) は "Omics annotation" の下位に位置し、上記の transcription data に対する意味づけを記述するためのモジュールで、疫学的情報、遺伝子発現情報、実験設定・解析手法、Array 情報などが記述できるようになっている。巨大なモジュールのため、各詳細構造は (別添) 図 6~8 を参照されたい。例えば遺伝子発現情報 (図 6, "differentially expressed genes") では発現情報、実験設定・解析手法、Array 情報 (図 8) を記述し、"Epidemiology transcription" (図 7) では関連する疫学的情報を記述することが可能である。

これらのモジュールを iCOs モデルに追加することで、3.3 で述べたように iCOD データベースにおける Molecular information の各項目 (Array information, Statistical method, Transcription data) がマッピング可能となっている。

5 臨床オミックスデータベースの格納における課題

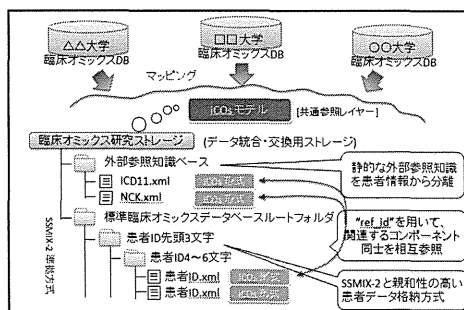
iCOs モデルは、(別添図 1, 2) に示すように主に (1) WHO-ICD11 や疾患知識データベース NCK (Normalized Clinical Knowledge) といった外部参照医学知識、(2) 各患者のオミックス情報、(3) 各患者の臨床情報の 3 つから構成されている。このうち、(2), (3) は (別添図 C) に示すように <OML> という「ある時点の患者オミックス/臨床情報」と、<Time Course of Clinical Information> という「患者の OML 時系列と予後情報」によって記述されるようになっている。このように iCOs 自体は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

6 解決方針の設定

5 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 D に示すような構造とすることにした。

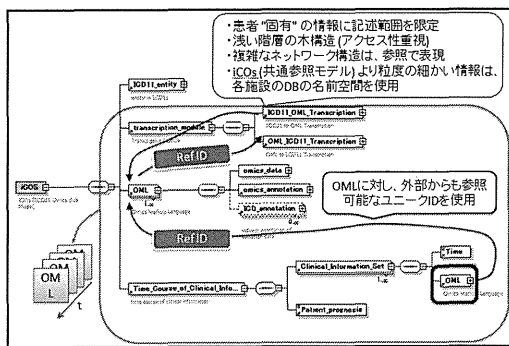


(図 D: データベース全体構造)

- 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

- 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 E のような XML 形式にすることとした。



(図 E) 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。

7 実データの格納

検証のため、6 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 23~28 に具体的な XML での記述例を示す。

【別添図 23】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 24】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 25】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相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 26】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タ、具体的な<Clinical Observation> などが格納されている。

【別添図 27, 28】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 27) 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 28) 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タが格納されている。

8 本方式で実現された内容

実データを用いた7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、”〇〇”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。本研究でモデル設計の原理として設定した“2-layer model”はこのようなデータベース統合の際に良く問題となる共通情報粒度の設定といった問題に対し、1つの解決策を与えるものであり、本研究の大きな特徴の1つである。既に国際標準としても確立しているGenomic sequence variation領域のモデルに加え、Transcriptomics領域の情報構造についても、既存のGO等の外部知識リソースの知見を基に分野専門家と検討し、共通参照レイヤーを制定した。臨床情報・オミックス情報・そしてICD11や疾患オントロジー(NCK)のような外部知識記述までを幅広くカバーしつつ、共通参照レイヤーによって多施設間情報連携を可能にしている臨床オミックス情報モデルは、世界的にもほぼ前例が無く、我が国発信の技術と

して非常に意義深いものである。

同様に、Time courseの記述モジュールは、元来静的な情報の記述を目的としたiCOsに対して、時系列情報の記述能力を与えるものであり、本研究の2つ目の大きな特徴である。臨床オミックス研究においては、時系列変化が重要な場合もあることから、今後より詳細なユースケースにおいて、さらなる検証を進める予定である。

一方、Genomic sequence variationならびにTranscriptomics以外のオミックス領域、例えばProteomics, Signalomics, Metabolomics, Organomics等の情報については、本年度は検証が行えておらず、今後の課題である。しかしこれらの領域はまだ発展途上な分野も多く、標準的な知識の記述について未だ十分な知見が得られていないことから、当面の間は標準モデルの構築が困難であると考えられる。

同様に、非構造化データの取り扱いも今後の課題である。例えば自然言語で記述された放射線読影所見のようなデータは、今回構造化することなく、“laboratory findings”の下位の“diagnostic imaging”オブジェクトに格納することとしているが、本来はデータマイニングに耐えるような構造化形式にて記述されることが望ましい。しかし、多施設間で情報を収集する際に人手で構造化する方式は非常にコストが高い。そのため、現在自然言語処理技術を用いた自動構造化インポートモジュールの開発を検討している。

次に、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と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。

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〇〇を発症した患者」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な粒度の情報項目は 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マ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

ICD11 臨床オミックス情報サブモデル (ICD11 Omics Sub information model: iCOs) の開発とその妥当性の検討は、オミックス医学を ICD 疾病分類に組み込むための要に位置づけられる研究であり、医学研究、臨床医療を含めた医学全体の発展に貢献することができる。さらに現実的な利用観点での実証を行った臨床オミックス情報モデルの提示は、網羅的オミックス情報の医学研究応用をデータ交換の効率性を向上という観点から加速することができる。

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。さらに、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。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

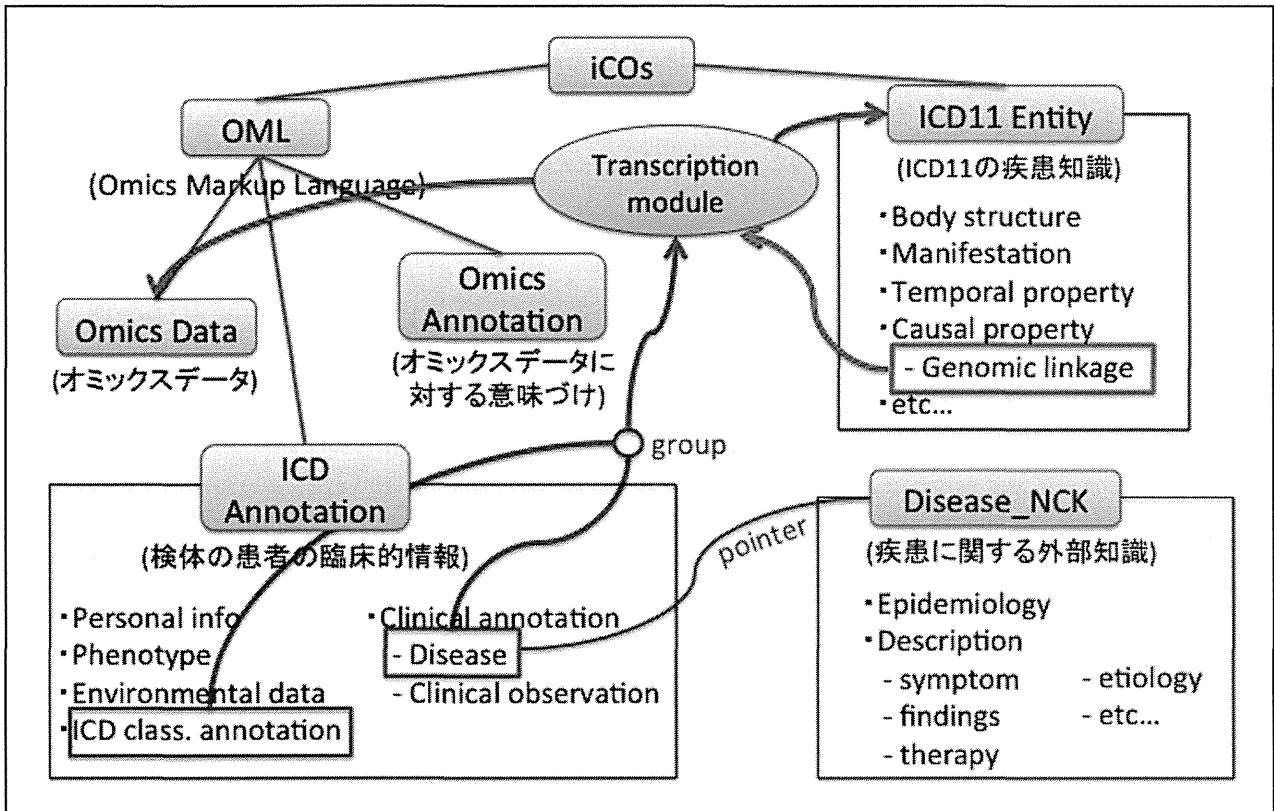
- 1-1. Hiroshi Tanaka, Masahiro Nishibori, Jun Nakaya. 2012. Disaster-Tolerant Architecture of Regional Healthcare System with Special Reference to Great Japan East Earthquake Disaster. 5th WSEAS International Conference on NATURAL HAZARDS (WSEAS2012- NAHA '12). 32-35.
- 1-2. Jun Nakaya, Teiji Tominaga, Hiroshi Tanaka, Masayuki Yamamoto. 2012. Disaster-tolerant features of ICT Architecture in Tohoku Medical Megabank. APRU. 32-33.
- 1-3. Kyoko Hirano, Takako Takai-Igarashi, Jun Nakaya, Hiroshi Tanaka. 2012.OCT. "An Interaction Network of Biological Entities Extracted from the Literatures on Aging Studies". JSBi 2012
- 1-4. Hashime Sawai, Takako Takai-Igarashi, Jun Nakaya, Hiroshi Tanaka. 2012.oct. "A Pathway Based Prioritization for Risk Genes of Bipolar Disorder". JSBi 2012
- 1-5. Tadashi Miyamoto, Soichi Ogishima, Jun Nakaya, Hiroshi Tanaka. 2012 Oct. "Expression trajectories" of reprogramming and differentiation on expression potential field. JSBi 2012
- 1-6. 中谷 純. 2012. 温泉の百科事典、ゲノム・ポストゲノム. 丸善. in Press.
- 1-7. 中谷 純、田中 博.(2012). 地域医療・福祉ネットワーク化白書. シードプランニング. 155-166
- 1-8. 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

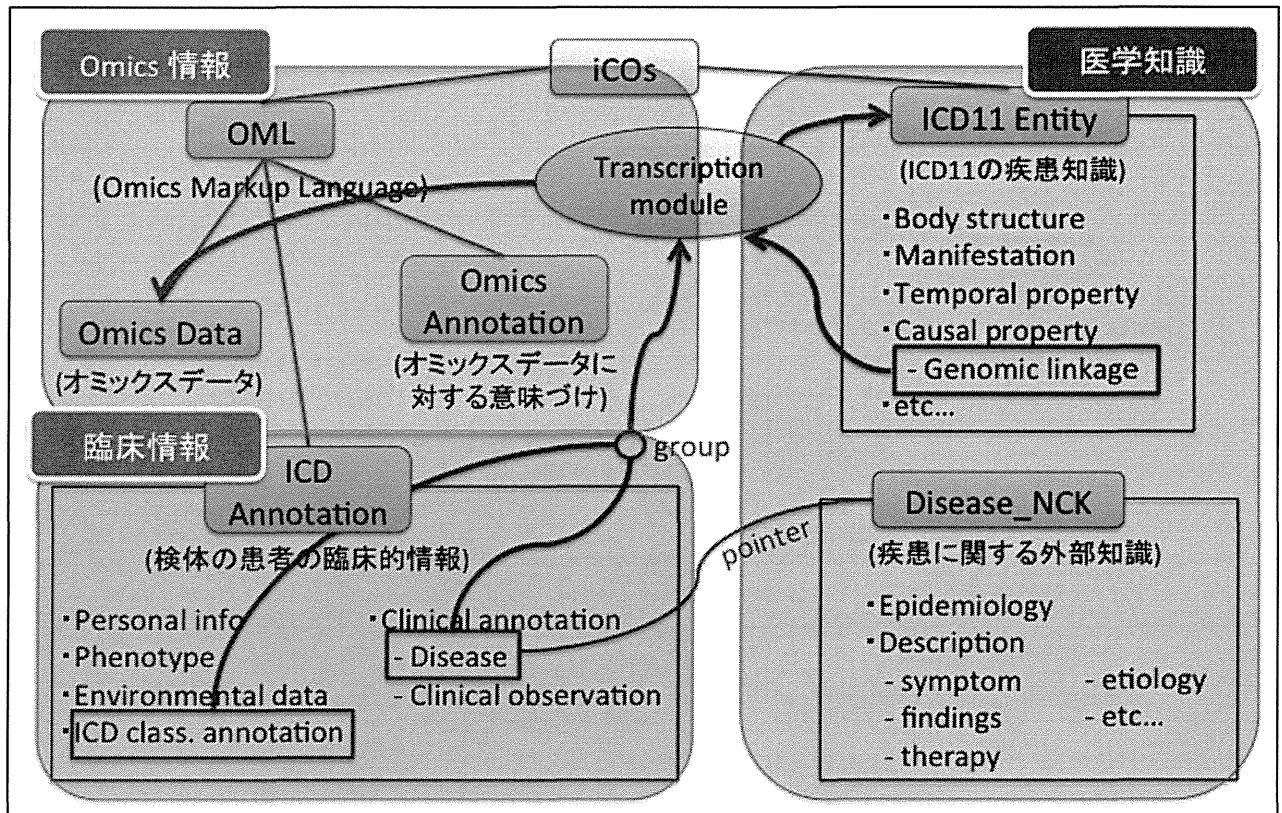
- 2-1. 医療情報と生命情報の接点 - 世界、日本、東北の現状と東北メディカル・メガバンク - 中谷 純
第 16 回日本医療情報学会春季学術大会 大会
企画 函館、2012.6.2
- 2-2. みやぎ医療福祉情報ネットワーク協議会(MMWIN)
と 総務省・東北地域医療情報連携基盤構築事業
中谷 純 総務省 ICT フェア 仙台、
2012.6.27
- 2-3. 「未来型医療の実現のために必要な医療情報基盤」
中谷 純 並列生物情報処理イニシアティブ
(IPAB) ～創薬・医療とビッグデータ～ 東京、
2012.12
- 2-4. 地域医療福祉情報連携のあるべき姿・・・災害に
強く平時に役立つシステムとは 中谷 純
"TOMODACHI Fund: 地域医療福祉情報連携に
関する教育講座 第 1 回 地域医療福祉情報連携の歩み
と考え方" 仙台、2012.10.25
- 2-5. "医療情報共有の意義と課題：未来型医療において
必要な社会情報基盤" 中谷 純 ゲノム研究の法的課題
についての研究会 上智大、2012.12.22
- 2-6. 未来型医療と医療情報 中谷 純 早稲田大学
第 10 回神戸研究者全体会議 神戸 T R I、
2012.11.30
- 2-7. 未来型医療 と 今後の医療情報基盤 中谷 純
宮城県官公立病院事務長会 松島、
2012.11.15
- 2-8. みやぎ医療福祉情報ネットワーク協議会(MMWIN)
と 総務省・東北地域医療情報連携基盤構築事業
中谷 純 全国病院会 仙台、2012.10.05
- 2-9. みやぎ医療福祉情報ネットワーク協議会(MMWIN)
が 構築する 地域医療福祉情報連携基盤 中谷 純
仙台市立病院講演会 仙台、2012.8.1
- 2-10. 災害における医療情報とその対策システム -
現実、ヒューマンネットワーク、実システム -
中谷 純 第 3 回日本医療情報学会秋季連
合大会 2012.11.17
- 2-11. 日経 B P Business P h a r m a
"東北メディカルメガバンク機構の中谷 純教
授に聞く
「日本で初の階層型医療情報ネットワークが立ち
上がる」" 2012.5.8
- 2-12. Data FM "みんなでつくる未来の
医療 [医療情報を ICT 化するってどんなこと?]"
2012.9.17
- 2-13. 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
- 2-14. 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
- 2-15. 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
- 2-16. 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
- 2-17. MMWIN について 中谷 純 総務省視察 仙台、
2013.5.31
- 2-18. 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
- 2-19. 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
- 2-20. 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
- 2-21. 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
- 2-22. 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
- 2-23. 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
- 2-24. MMWIN について 中谷 純 総務省視察 仙台、
2013.7.29
- 2-25. Jun Nakaya. Essential Medical ITC Infrastructure
in Future Medicine. Medinfo2013.
- 2-26. 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
- 2-27. 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
- 2-28. 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
- 2-29. 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
- 2-30. 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
- 2-31. 宮城における地域医療介護情報連携 中谷 純
第 40 回日本産婦人科医学会学術集会 仙台、
2013.10.13
- 2-32. 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
- 2-33. 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
- 2-34. 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
- 2-35. 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
- 2-36. 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
- 2-37. 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
- 2-38. 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
- 2-39. 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
- 2-40. 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
- 2-41. 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
- 2-42. 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
- 2-43. 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
- 2-44. 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
- 2-45. 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
- 2-46. 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
- 2-47. 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
1. 特許取得 なし
- 以 上

(別添資料)



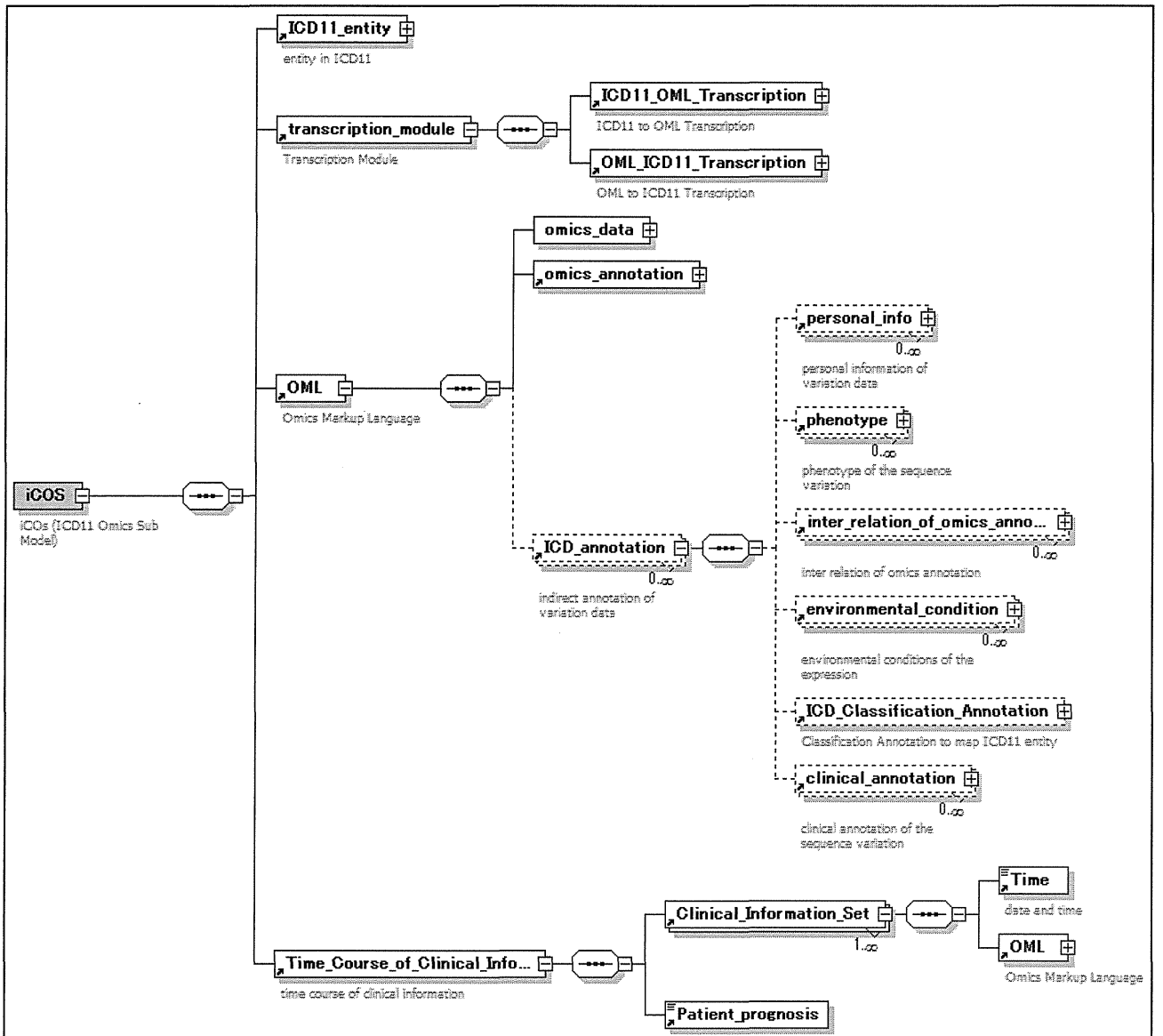
(図 1. iCOs モデル概要)



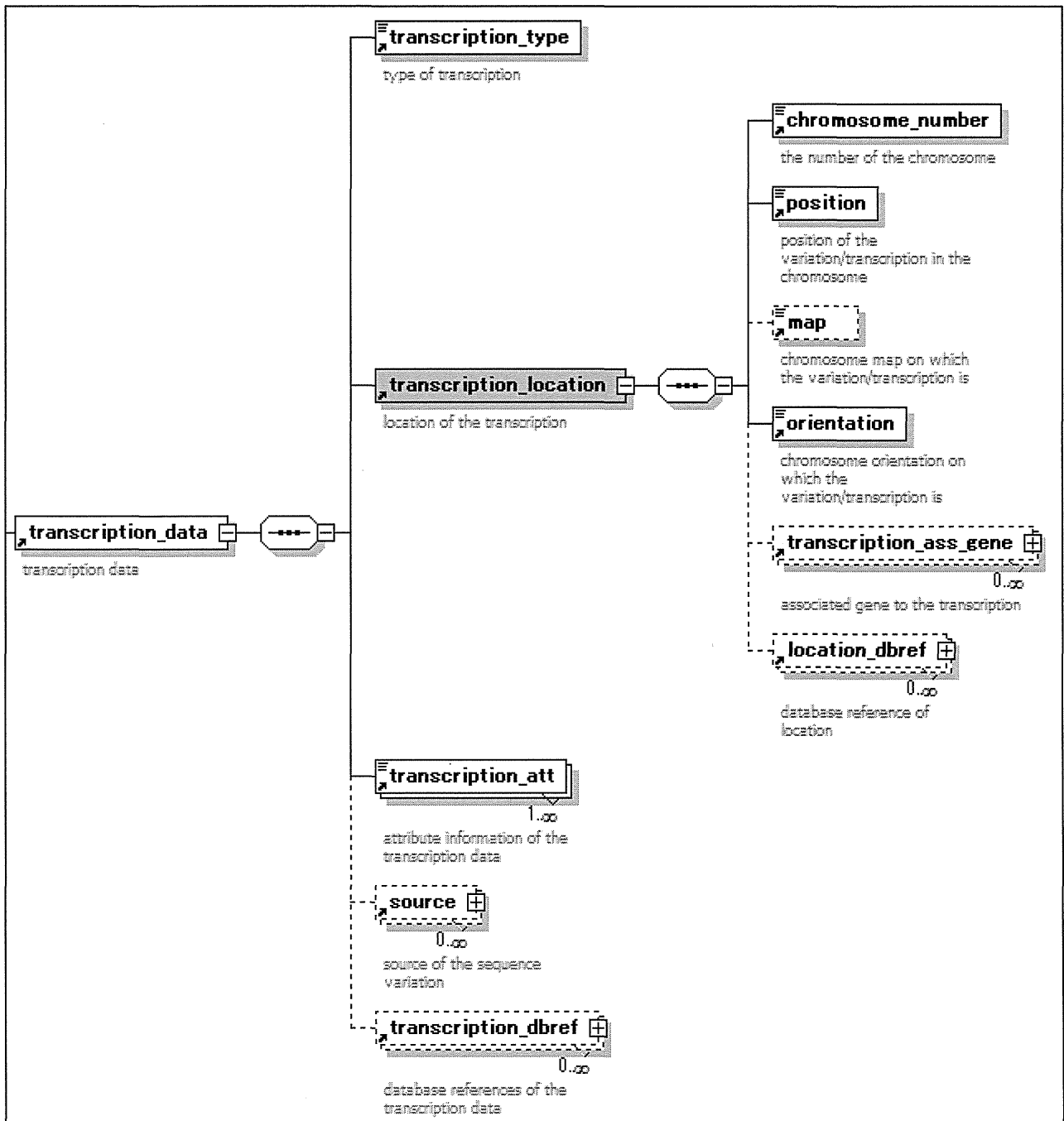
(図 2. iCOs モデルにおける記述情報種別)

		iCOD (網羅的疾患分子病態データベース)			iCOs (臨床オミックス情報モデル)									
		階層1	階層2	階層3	階層1	階層2	階層3	階層4	階層5	階層6	階層7	階層8		
1	Diagnosis	Diagnosis			iCOs	OML	ICD annotation	clinical annotation	disease NCK					
2	Primary Information	Primary Information	Demographic information	Age Sex						personal info	personal dbref			
3	Symptom	Symptom										subjective findings	symptoms	
4	Therapeutic Intervention	Therapeutic Intervention	Medical Intervention	Drug						clinical annotation	clinical observation	therapy	conservative	pharmaceutical
			Radiographic Intervention						radiation					
			Surgical Intervention						surgery					
5	Laboratory result	Blood test								objective findings	laboratory findings		blood test	
		Diagnostic Imaging							diagnostic Imaging					
		Endoscopic Screening							endoscopic Screening					
		Physique			physical findings									
6	Molecular Information	Array Information				omics annotation	transcription experiment analysis	differentially expressed gene (DEG) analysis	array information					
		Statistical Methods							DEG statistical analysis					
		Transcription Data							differentially expressed gene					
7	Pathology	Pathology	Macroscopic finding			ICD annotation	clinical annotation	clinical observation	objective findings	pathological findings	macroscopic finding			
			Tissue finding								tissue finding			
8	Patient Background	Medical History							medical history					
		Occupation Diet and Lifestyle							life history					
9	Prognosis	Prognosis				Time Course of Clinical Information	patient prognosis							

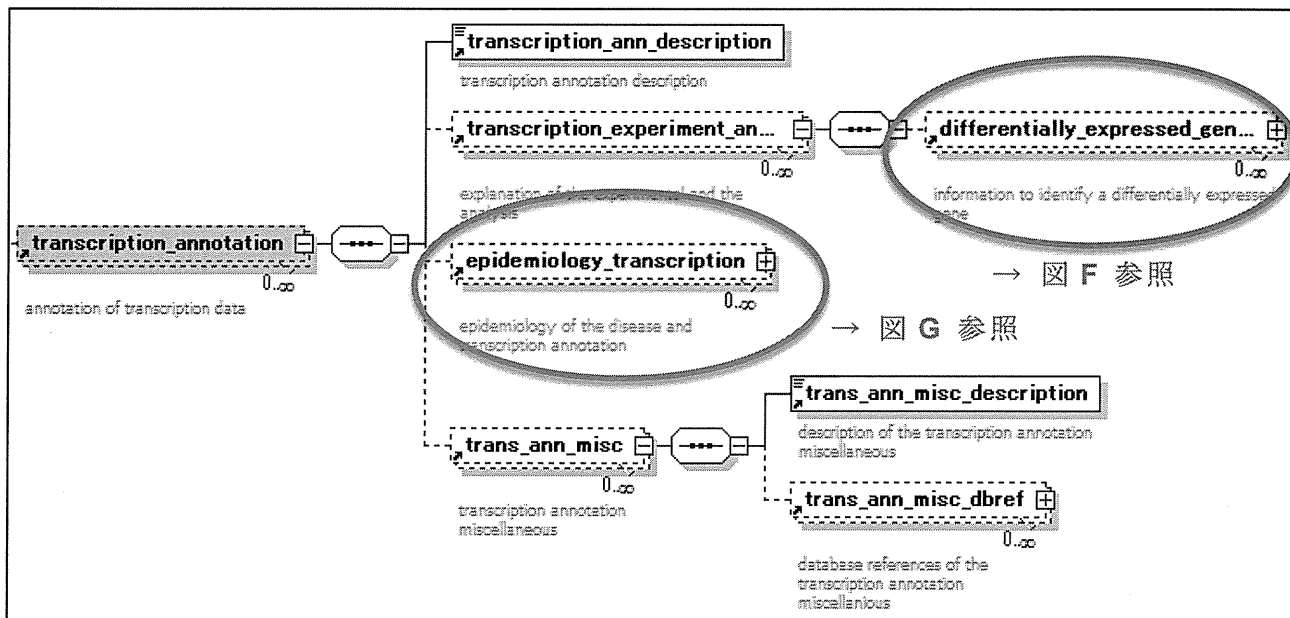
(表 1. iCOD と iCOs のマッピング結果)



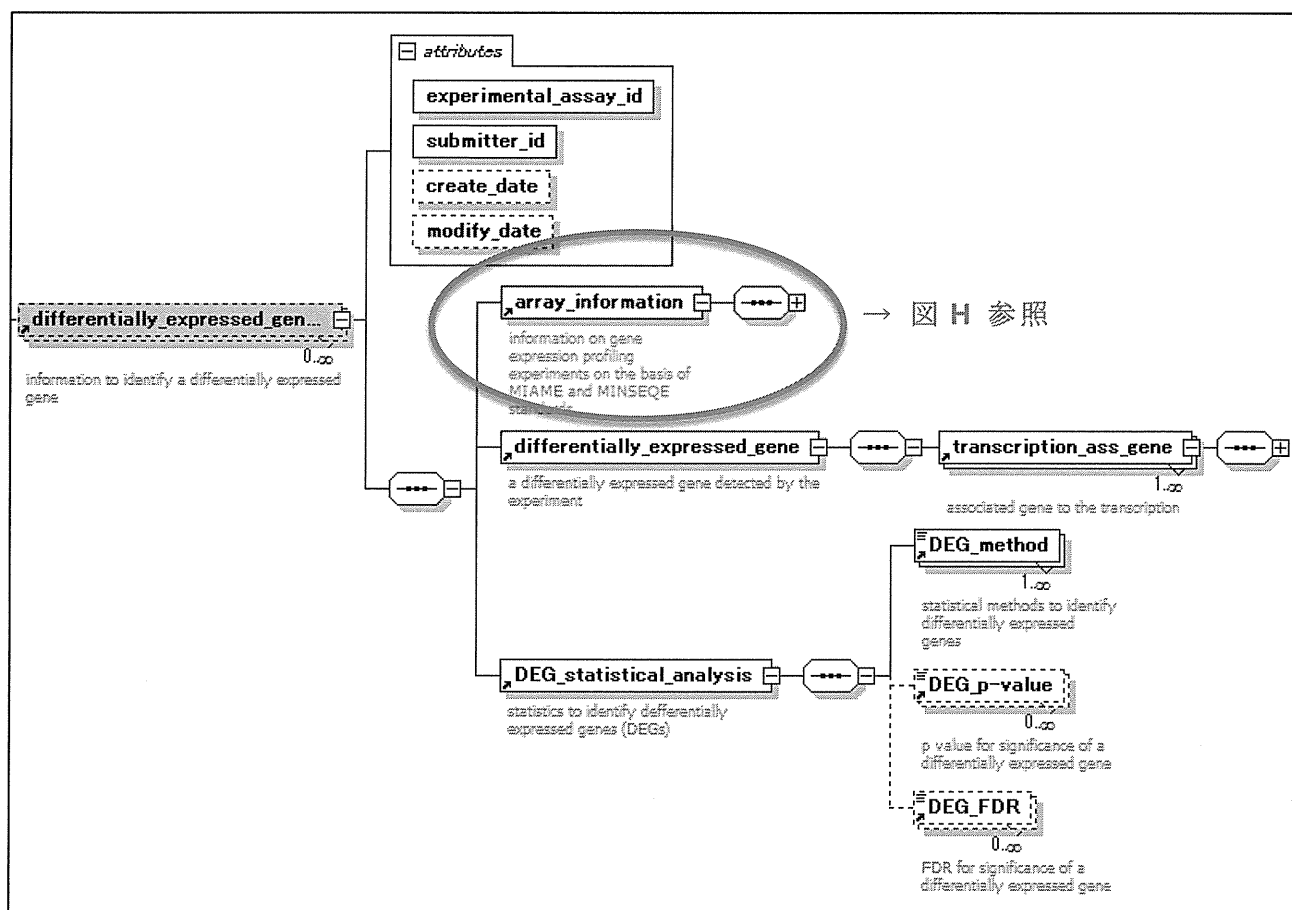
(図 3. iCOS モデルの上位構造と “Time course of clinical information” の位置づけ)



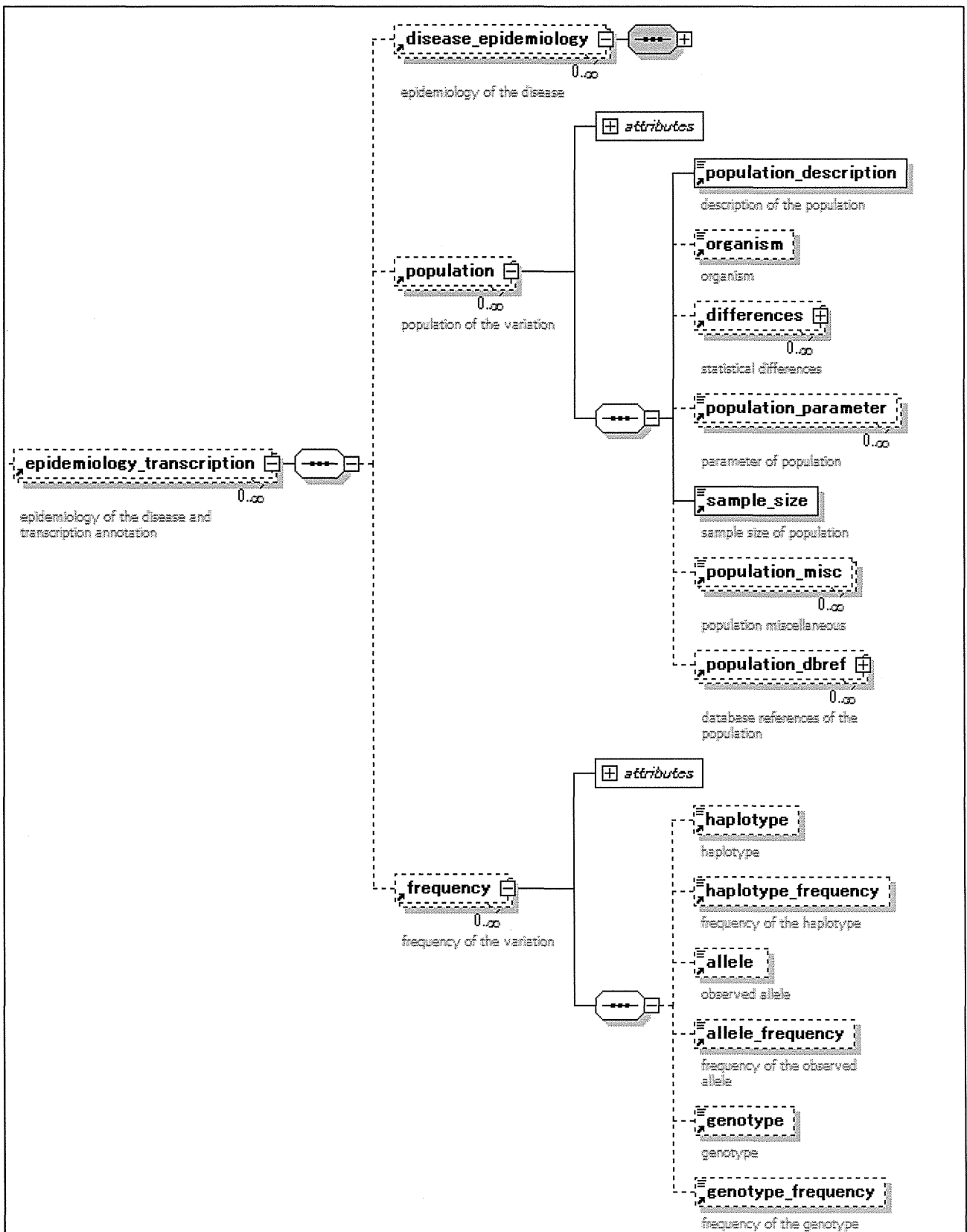
(図 4. Transcription_Data の詳細構造 (2 階層まで展開))



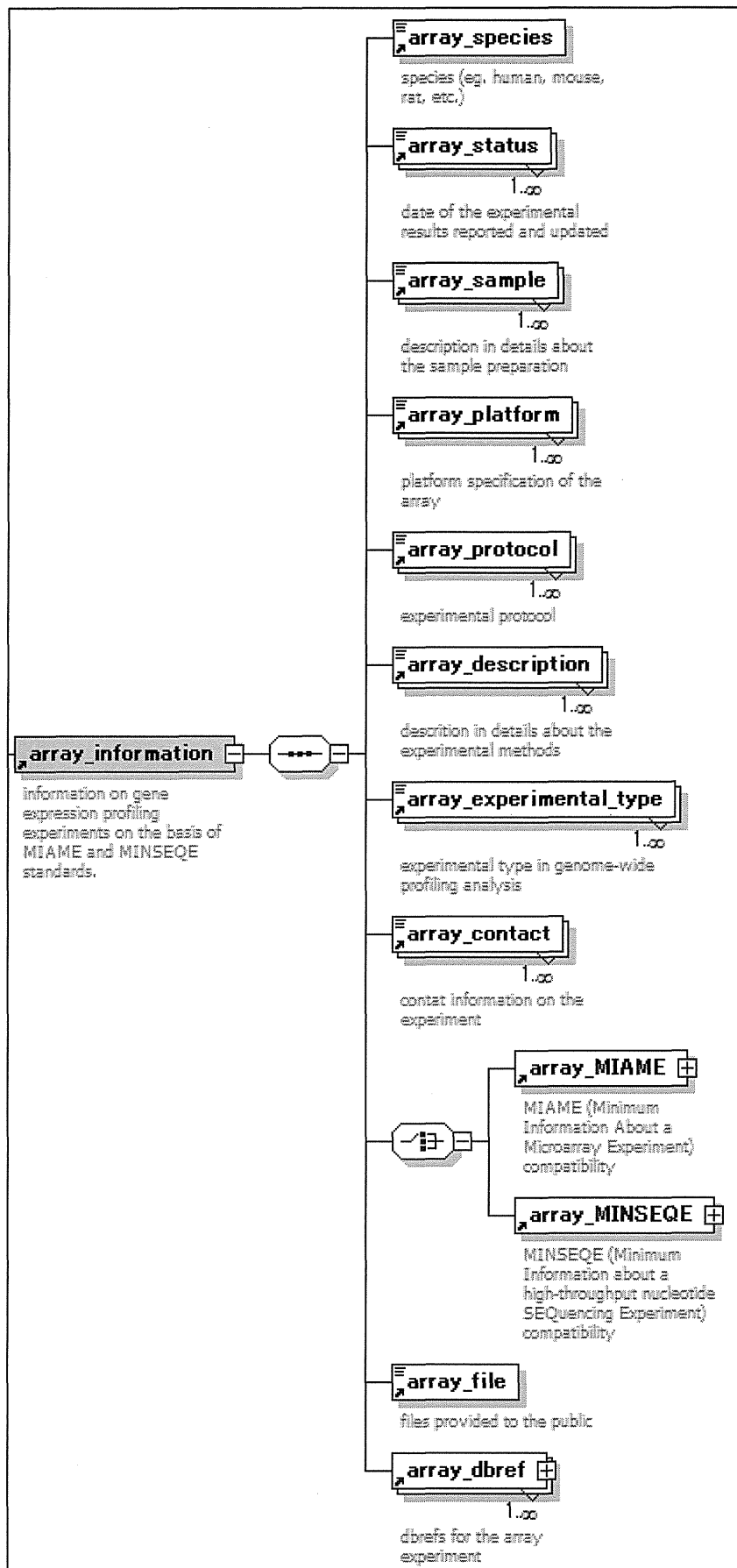
(図 5. Transcription_Annotation の詳細構造 (2 階層まで展開))



(図 6. "differentially_expressed_gene_analysis" の詳細構造 (2 階層まで展開))

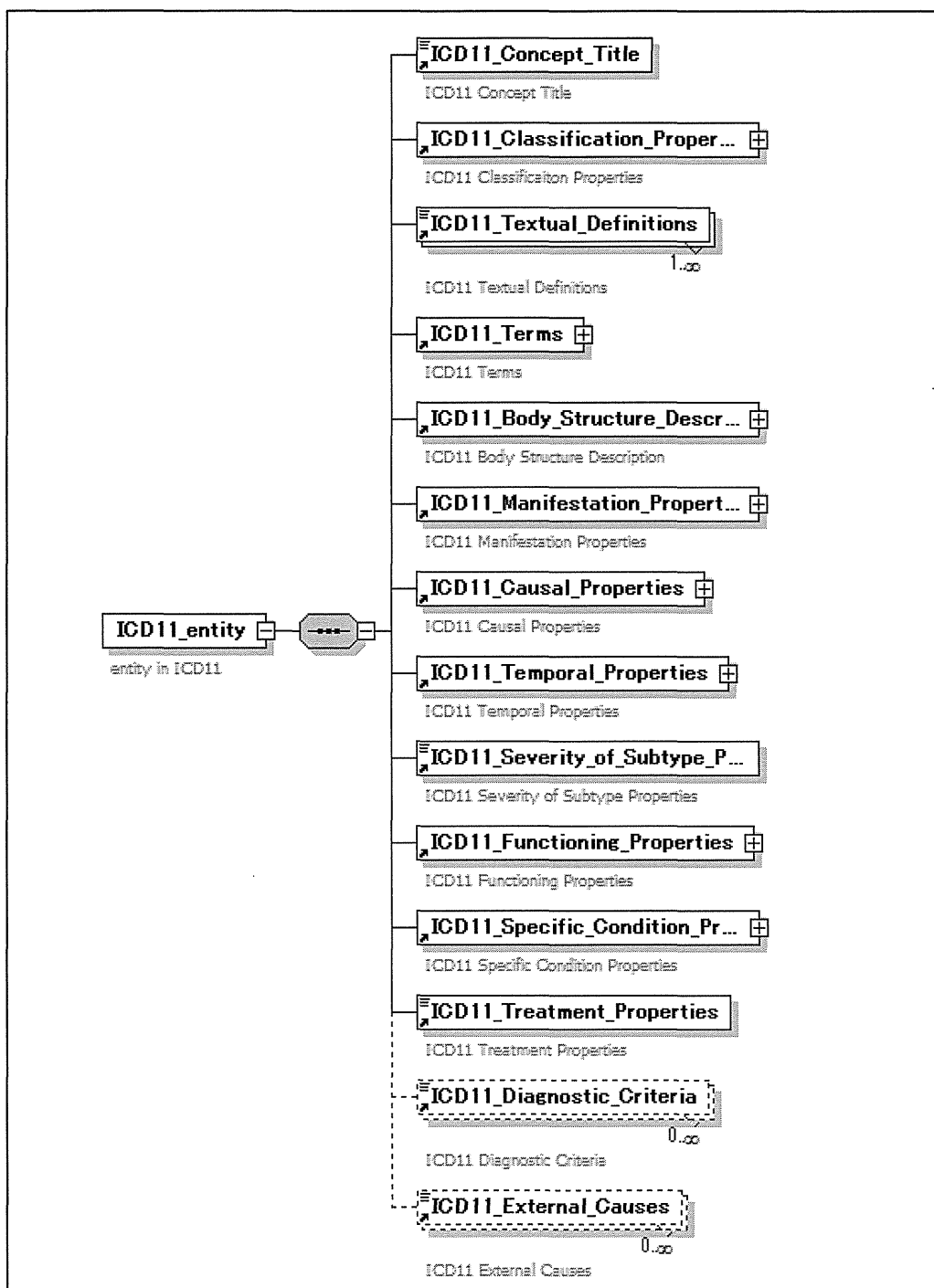


(図 7. “epidemiology_transcription” の詳細構造 (2 階層まで展開))



(図 8. “array_information” の詳細構造)

(以下は、iCOs モデルのうち、本年度で変更を加えなかった部分についての参考資料)



(図 9. ICD11_entity の概要 (第 1 階層まで展開))