

201202006A

厚生労働科学研究費補助金
政策科学総合研究事業（統計情報総合研究）

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 総括研究報告書

研究代表者 中谷 純
(東北大学)

平成25（2013）年 5月

目 次

I. 総括研究報告

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 中谷 純
-----1

II. 研究成果の刊行に関する一覧表 ----- 23

平成 24 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ことである。本年度においては、ユースケースの設定、並びに iCOD 等の実データを用いた iCOs モデルの記述可能性検討を行い、結果に基づいて iCOs モデルの修正を行った。いくつかの拡張を行うことで、iCOs モデルが臨床オミックス実データを十分に記述することが可能であることが示された。

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 臨床オミックス情報サブモデル (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臨床オミックス情報サブモデル (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、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を目的とする。

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)が包含されているが、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

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

野専門家の協力の下に行った。

最後に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。

本研究では、患者情報を扱わないため特に倫理面での配慮は必要としない。

C. 研究結果

以下、主要な結果の概要について述べる。iCOs の詳細や、マッピング結果の詳細については、別添の図表ならびに XML Schema を参照されたい。

C-1 ユースケースの設定

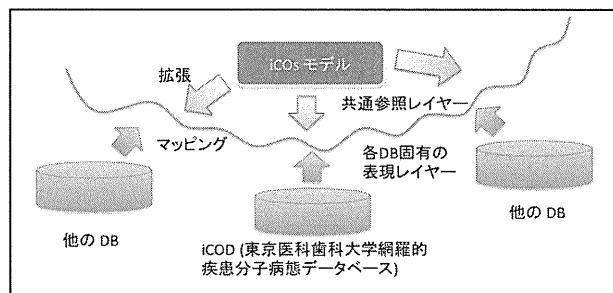
まず iCOs モデルがどうあるべきかを考える上でユースケースの設定は必要不可欠である。本年度は以下のようなユースケース概要を設定し、全体設計を行っていくこととした。

- 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。
- 臨床オミックス研究データベースを新規に構築する際の設計指針を提供する。

C-2 “Mapping principle” の設定

C-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1)。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



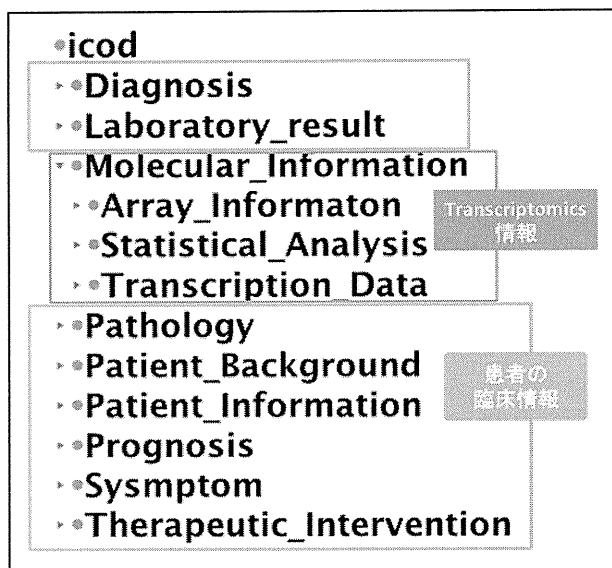
(図1) 2-layer モデルの考え方

C-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。

C-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。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”と、その検体の患者の臨床情報である



(図 2) iCOD データベース情報項目 (上位)

“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 モデルで記述するという事は、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事に他ならない。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。

C-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 A に示す。

iCOs で記述される情報は (別添) 図 B 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 C~R を参照されたい。

C-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 に格納する際には、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 A 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

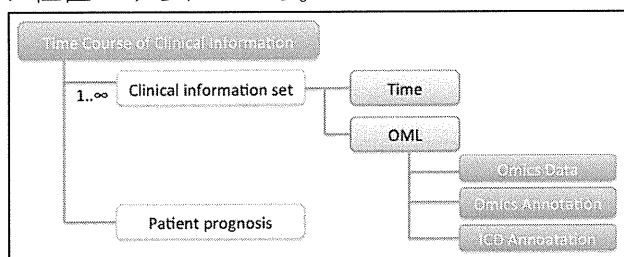
ルの考え方を大きく変更することなく、少数の項目を追加するだけで iCOD データベース項目 (共通参照レイヤーの情報粒度) とのマッピングが可能であることが判明した。

C-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。

C-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 3 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 3. Time course of clinical information)

C-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 D~H に示すように記述した。

“Transcription data” (別添 図 D) 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 E) 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 F~H を参照されたい。例えば遺伝子発現情報 (図 F, “differentially expressed genes”) では発現情報、実験設定・解析手法、Array 情報 (図 H) を記述し、“Epidemiology transcription” (図 G) 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、C-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。

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。今年度は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、本年度追加した 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。

また、今回は東京医科歯科大学の iCOD データベースでの検証を行ったが、多施設での相互運用性についても実データを用いた検証が望ましい。その際にはデータの複雑性への対応、大量のデータの取り扱い、など複数の観点での検証が重要であると考えられる。これについても次年度以降引き続き検討していく予定である。

E. 結論

本年度は、ICD11 臨床オミックス情報サブモデル (ICD11 Omics Sub information model: iCOs) の実用性を多角的に検証するため、実際の臨床オミックスデータベース (東京医科歯科大学網羅的疾患分子病態データベース iCOD) や、GO, SNOMED-CT などの他の実オントロジーデータを用いて、iCOs モデルによる記述可能性並びに、モデルの修正点について調査を行

った。結果、iCOs モデルの基本的な考え方に大きな変更を加えることなく、所見に関する詳細項目の追加、診療録の時系列記述モジュール、Transcription に関する記述モジュールを追加することで、臨床オミックス実データの記述が行えることが確認された。現在、2015 年のリリースを目指し ICD11 の改訂作業が進められているが、ICD11 のコンテンツモデルに組み込みが可能で、疾患知識と臨床情報ならびにオミックス情報を横断的に記述することができる iCOs モデルは、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。次年度以降ではこの iCOs モデルをより強固なものとするため、iCOs モデルで表現された iCOD の実データベースデータを計算機処理する上での複雑性対応・多量データ対応・多視点対応能力について検討を進めると共に、異なる施設間 (東京医科歯科大学・東京大学を予定) での相互運用性についての検証を行う予定である。

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

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

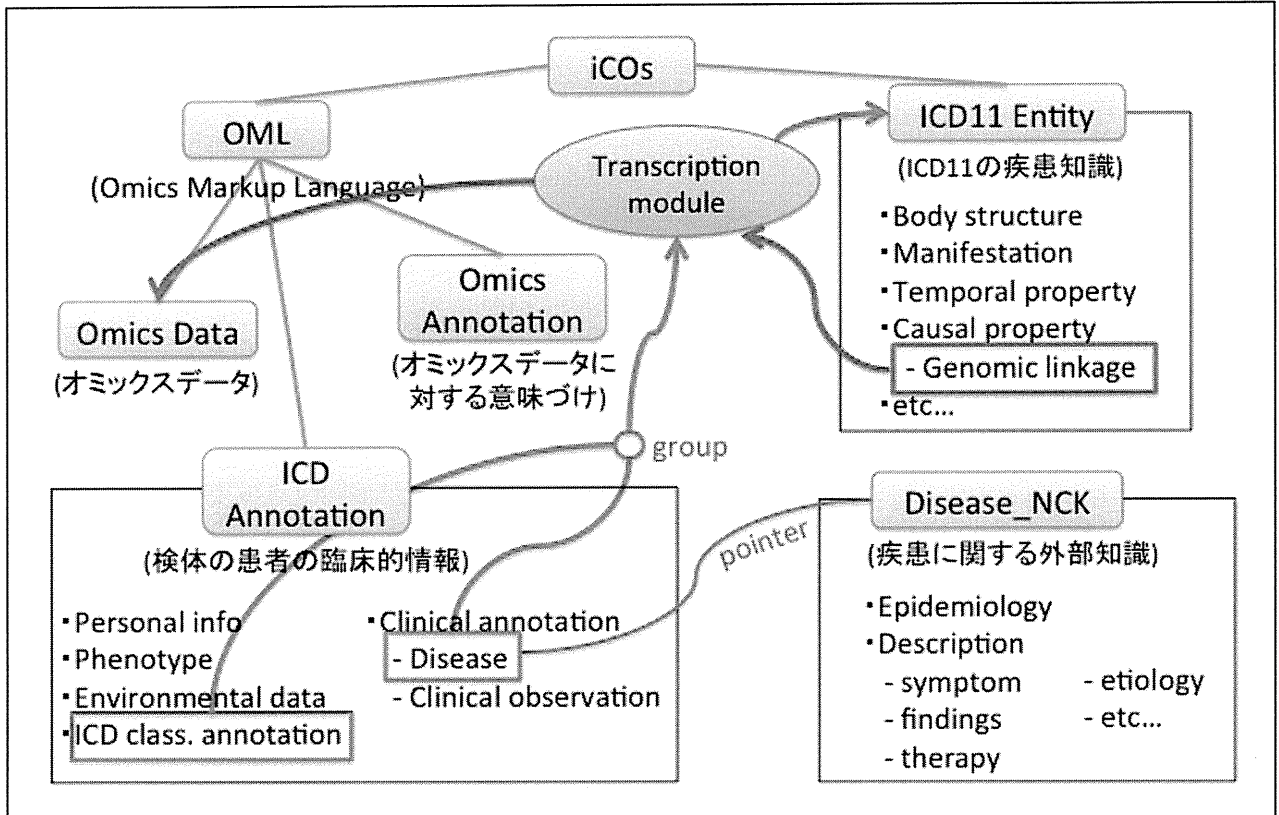
2-10. 災害における医療情報とその対策システムー 現実、ヒューマンネットワーク、実システムー 中谷 純 第 3 2 回日本医療情報学会秋季連合大会 2012.11.17

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

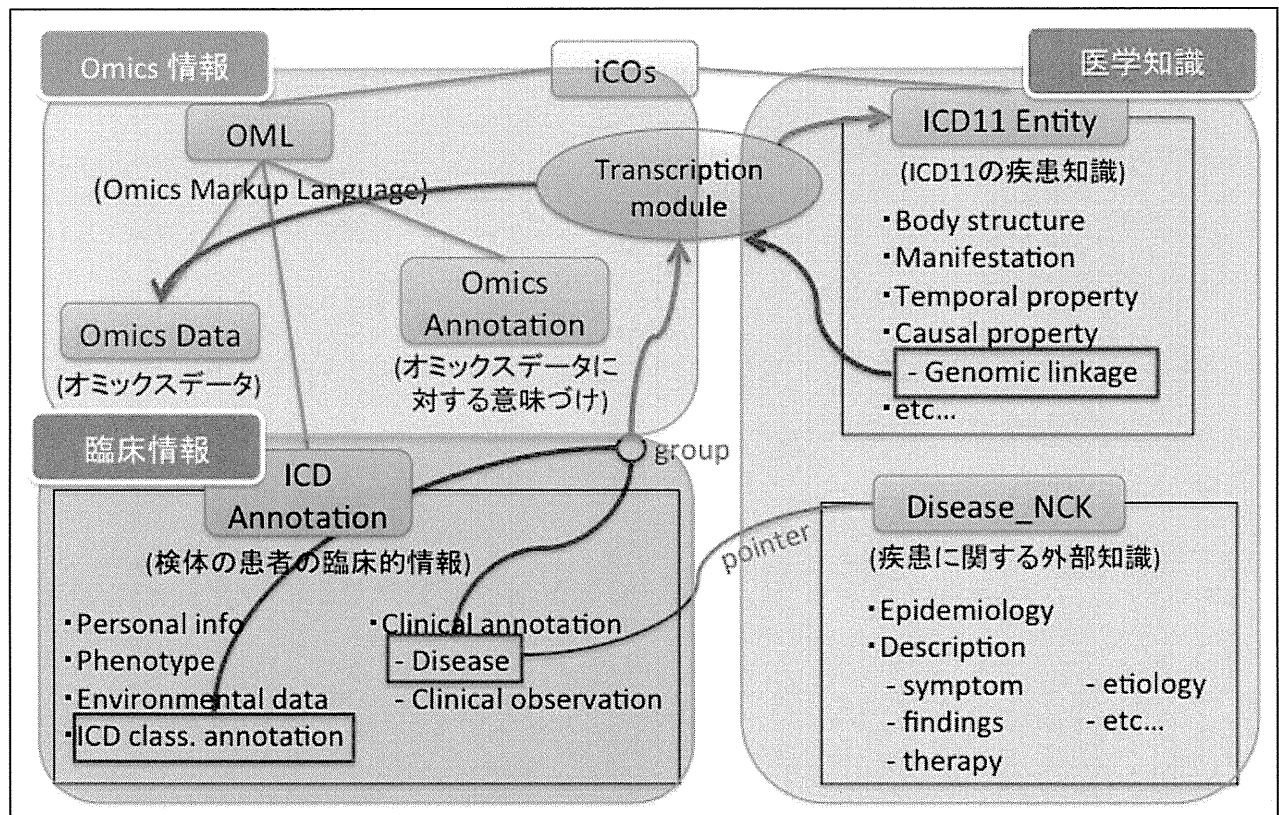
2-12. D a t a F M "みんなでつくる未来の医療 [医療情報を ICT 化するってどんなこと?]" 2012.9.17

H. 知的財産権の出願・登録状況

なし



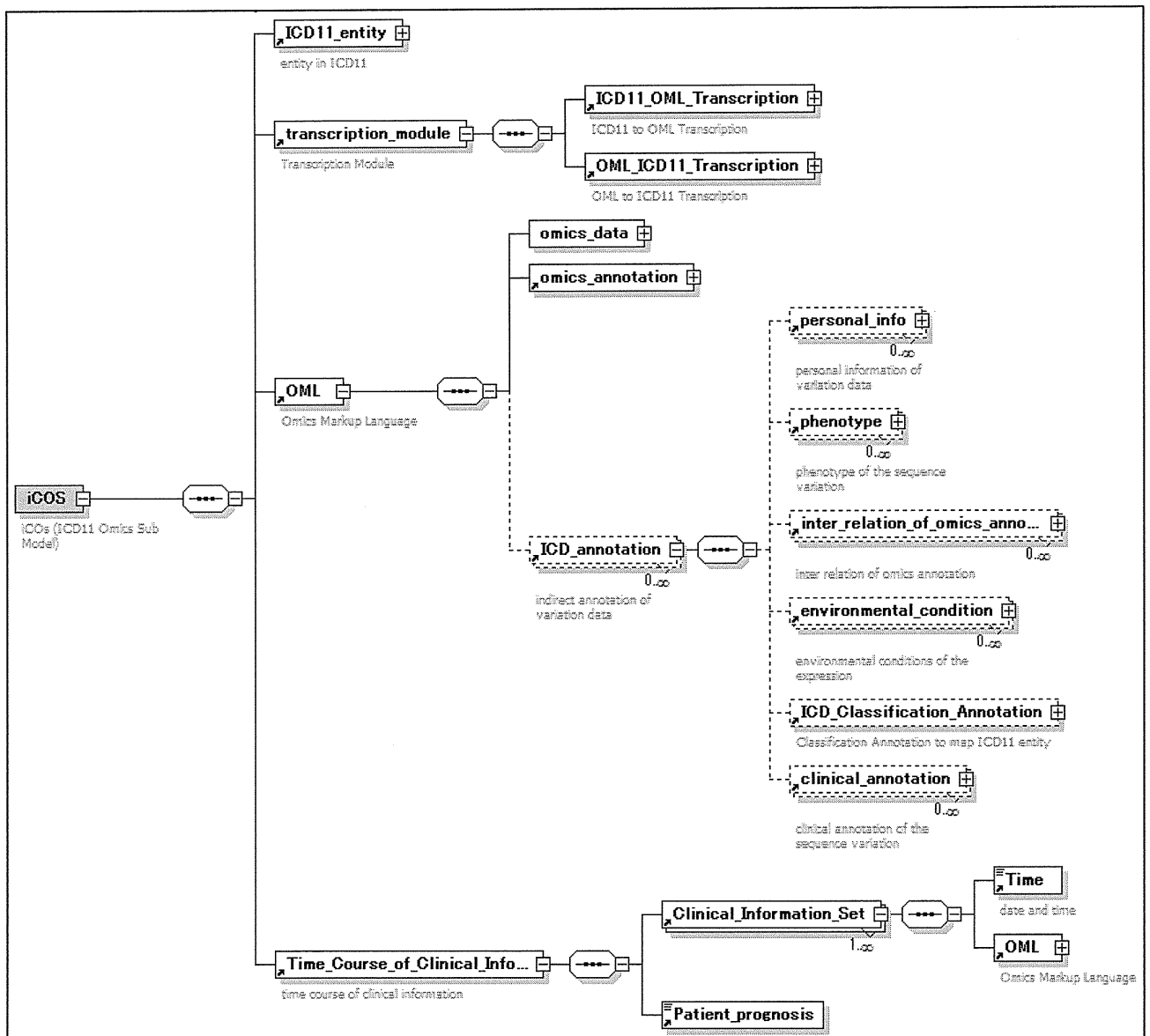
(図 A. iCOs モデル概要)



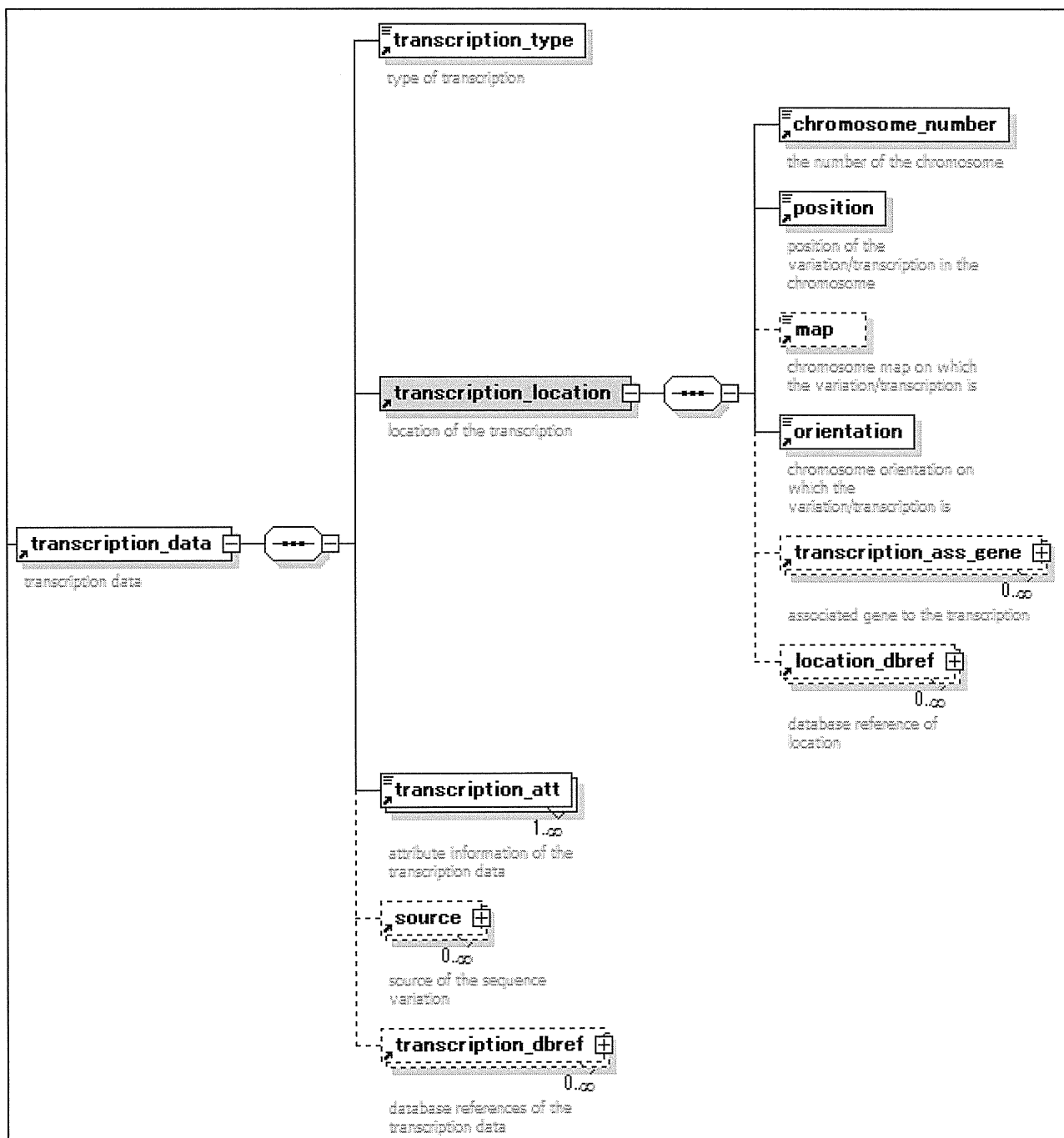
(図 B. iCOs モデルにおける記述情報種別)

	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							

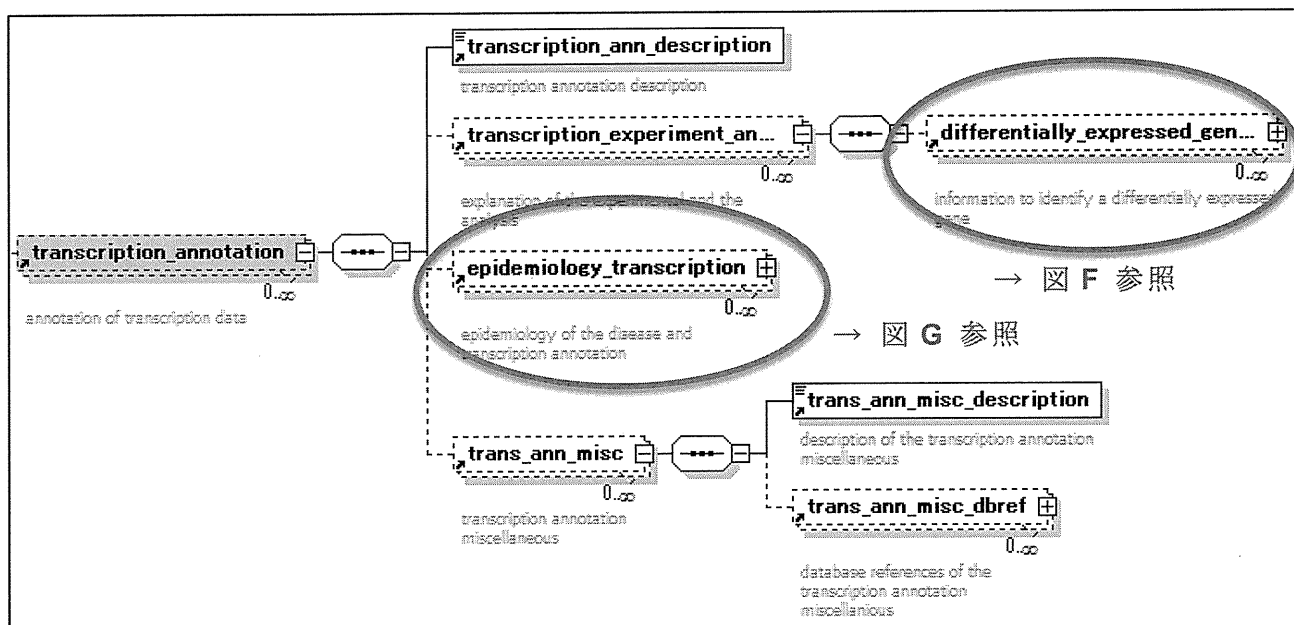
(表 A. iCOD と iCOs のマッピング結果)



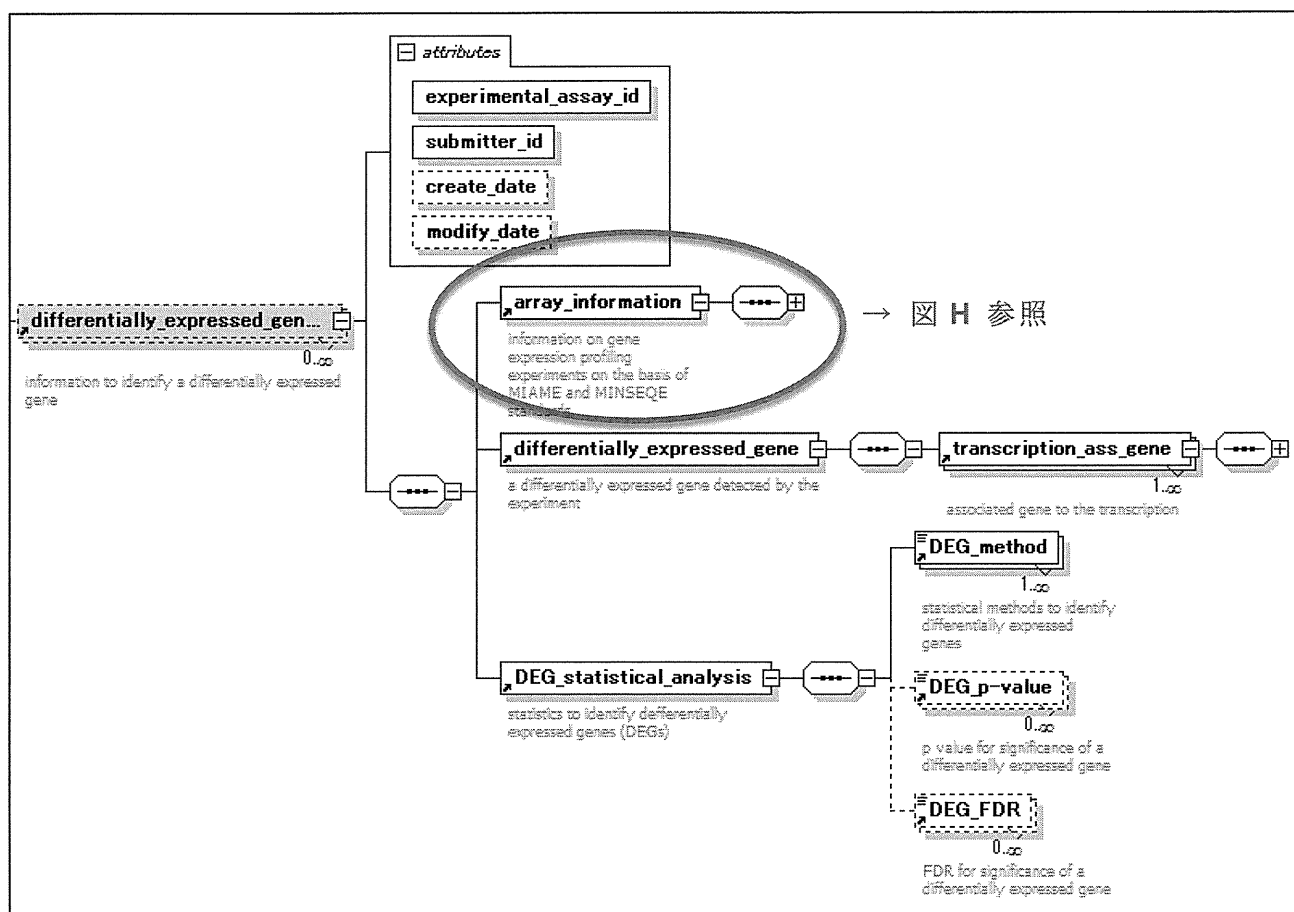
(図 C. iCOS モデルの上位構造と “Time course of clinical information” の位置づけ)



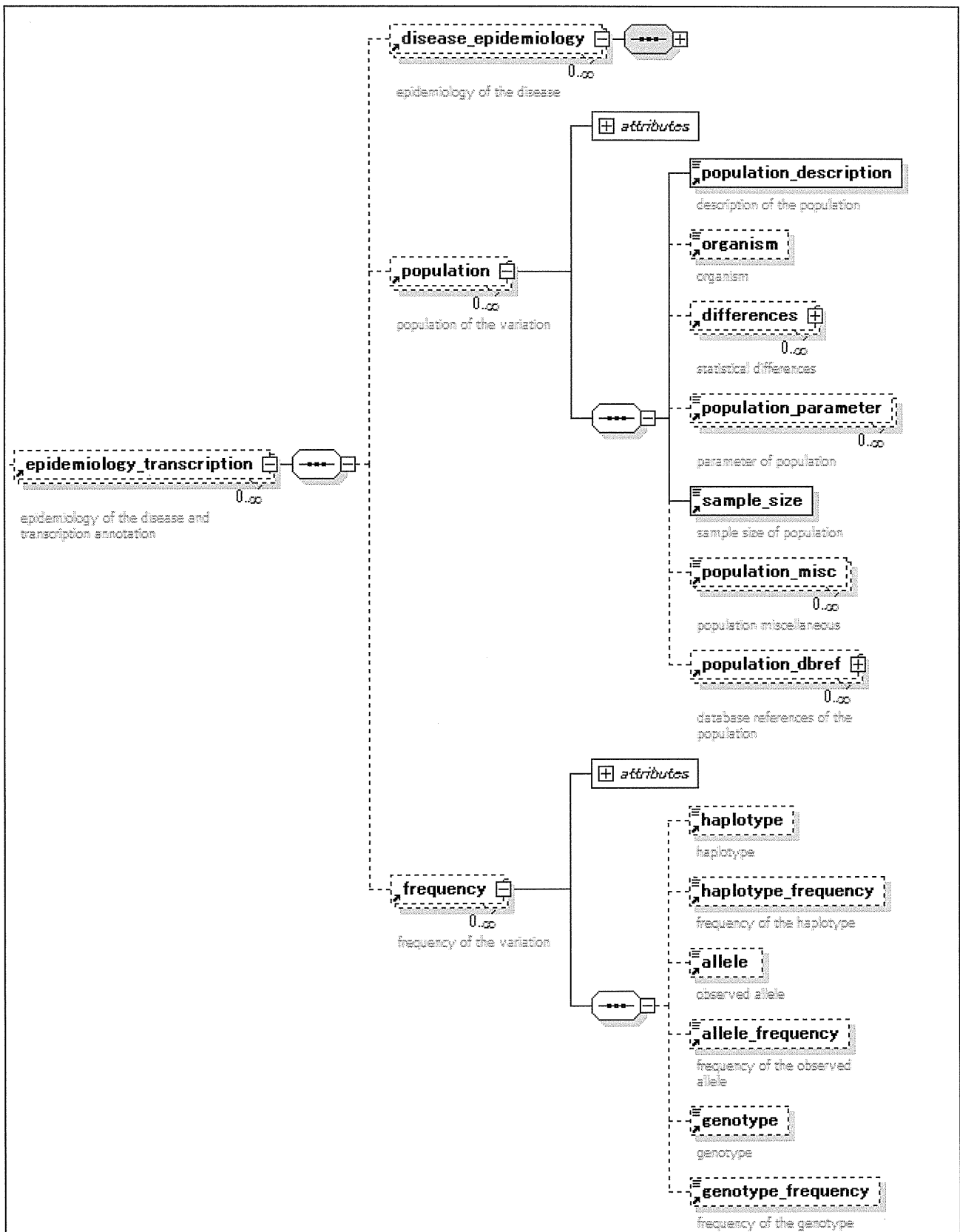
(図 D. Transcription_Data の詳細構造 (2 階層まで展開))



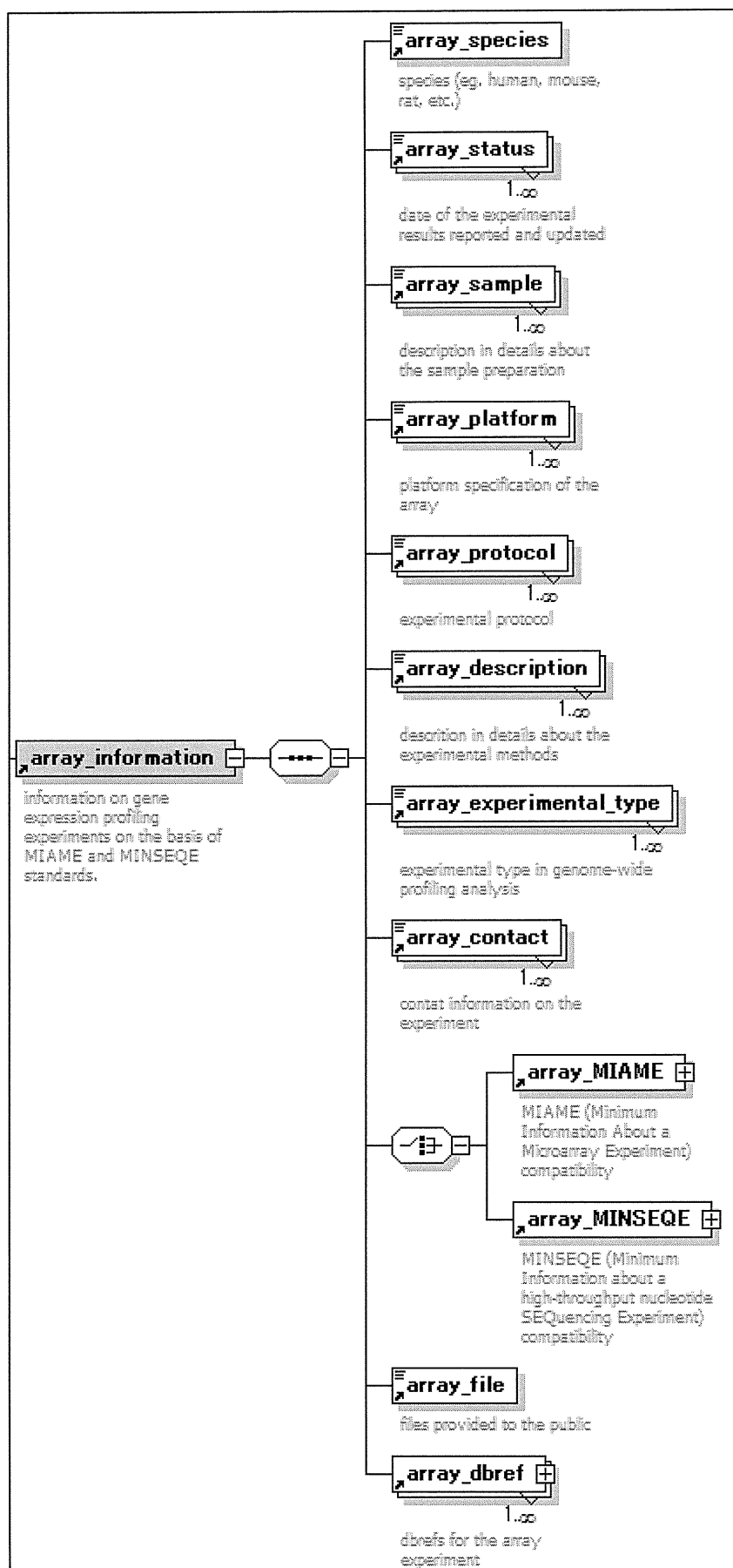
(図 E. Transcription_Annotation の詳細構造 (2階層まで展開))



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

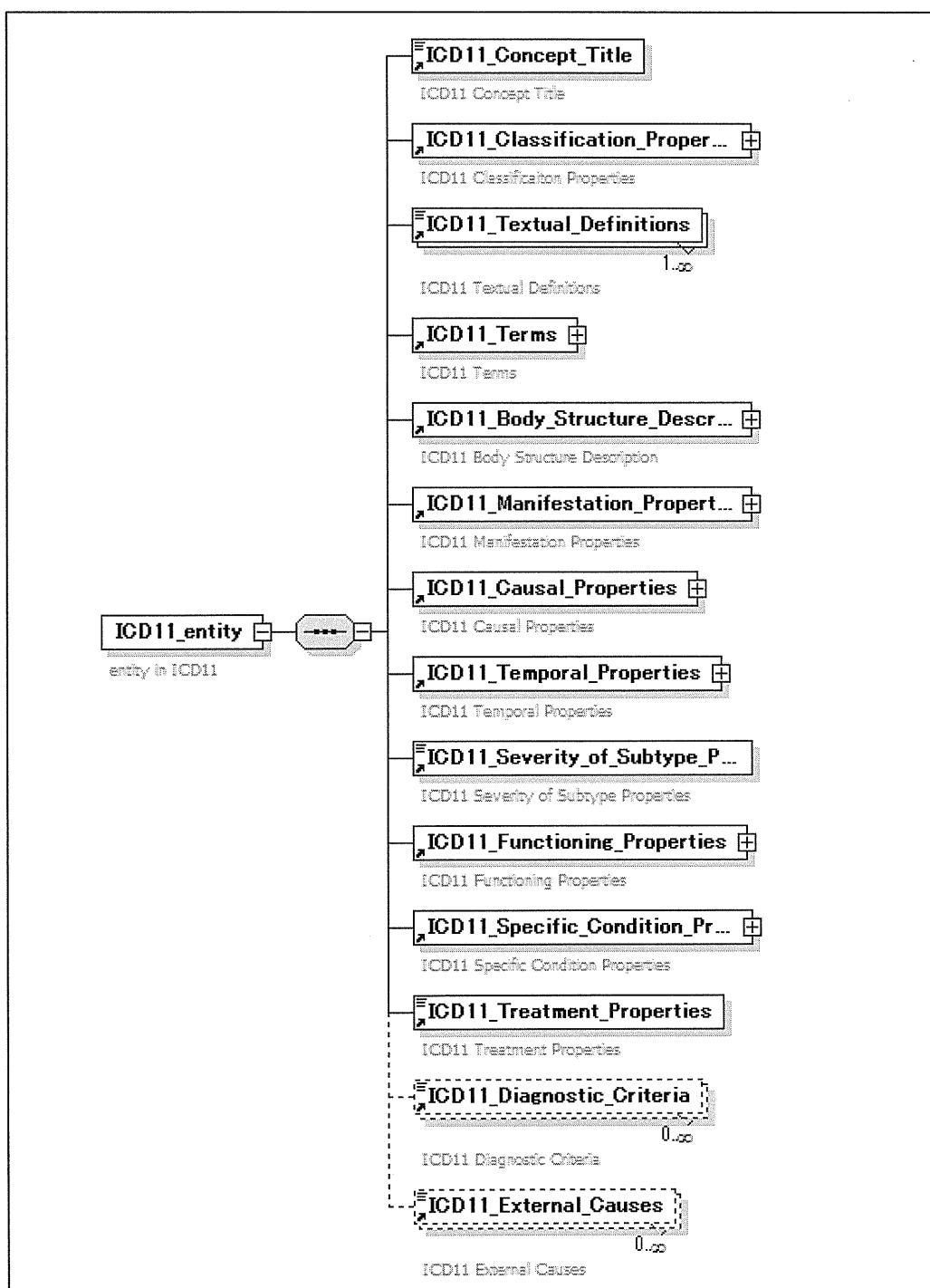


(図 G “epidemiology_transcription” の詳細構造 (2 階層まで展開))

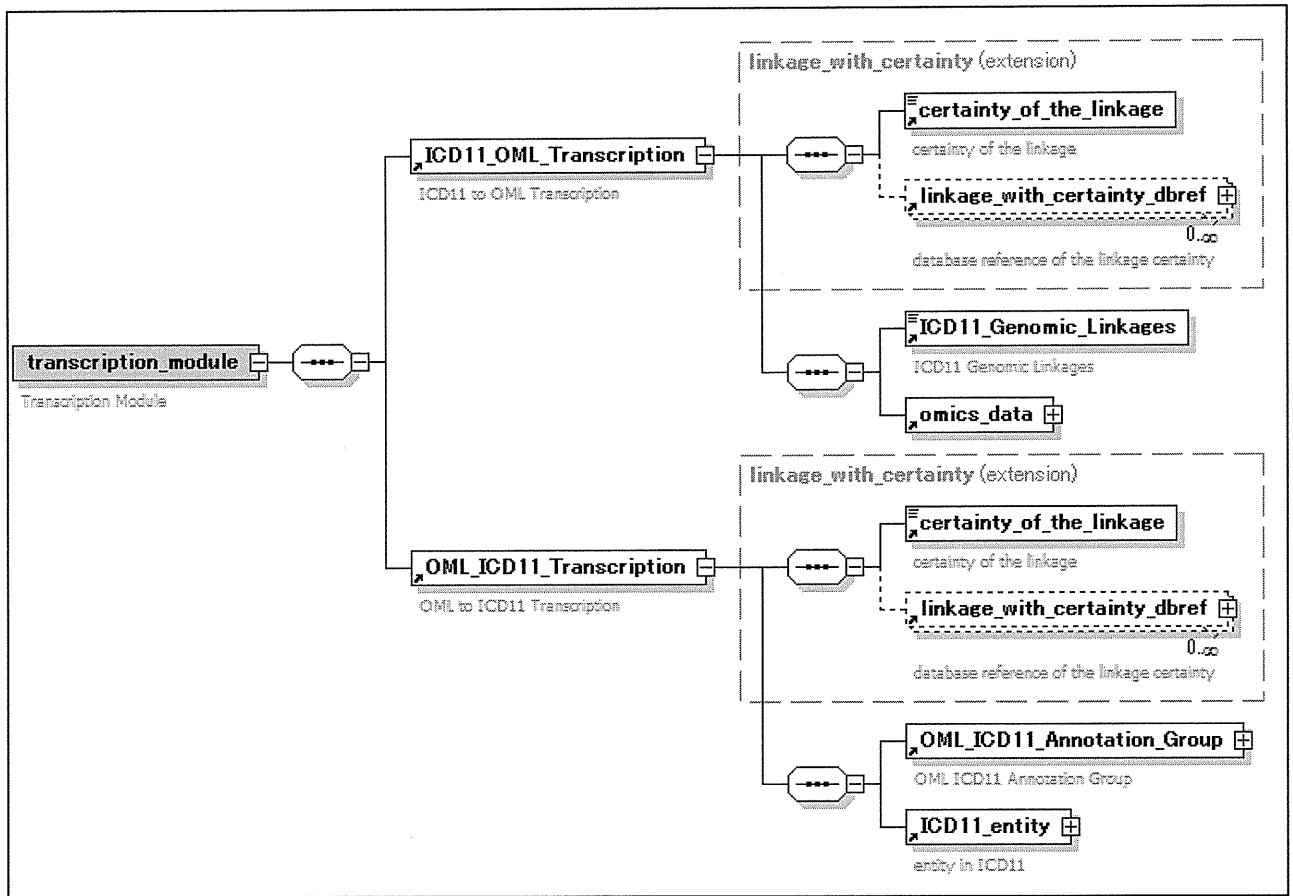


(図 H. “array_information” の詳細構造)

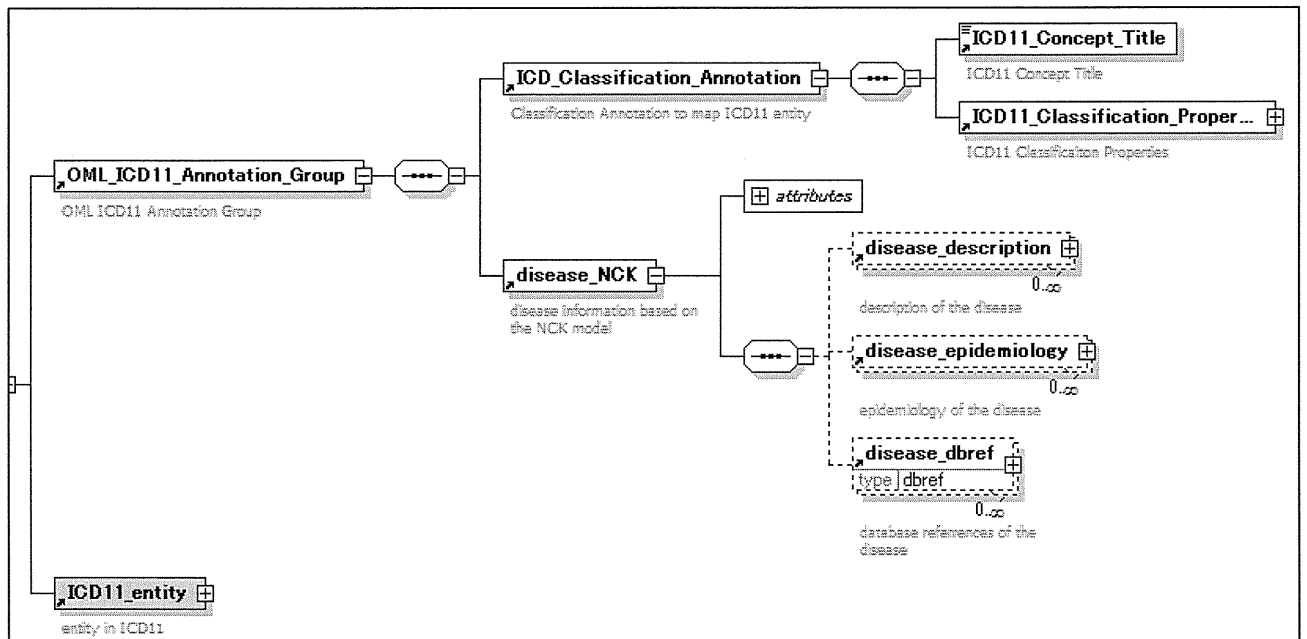
(以下は、iCOs モデルのうち、本年度で変更を加えなかった部分についての参考資料)



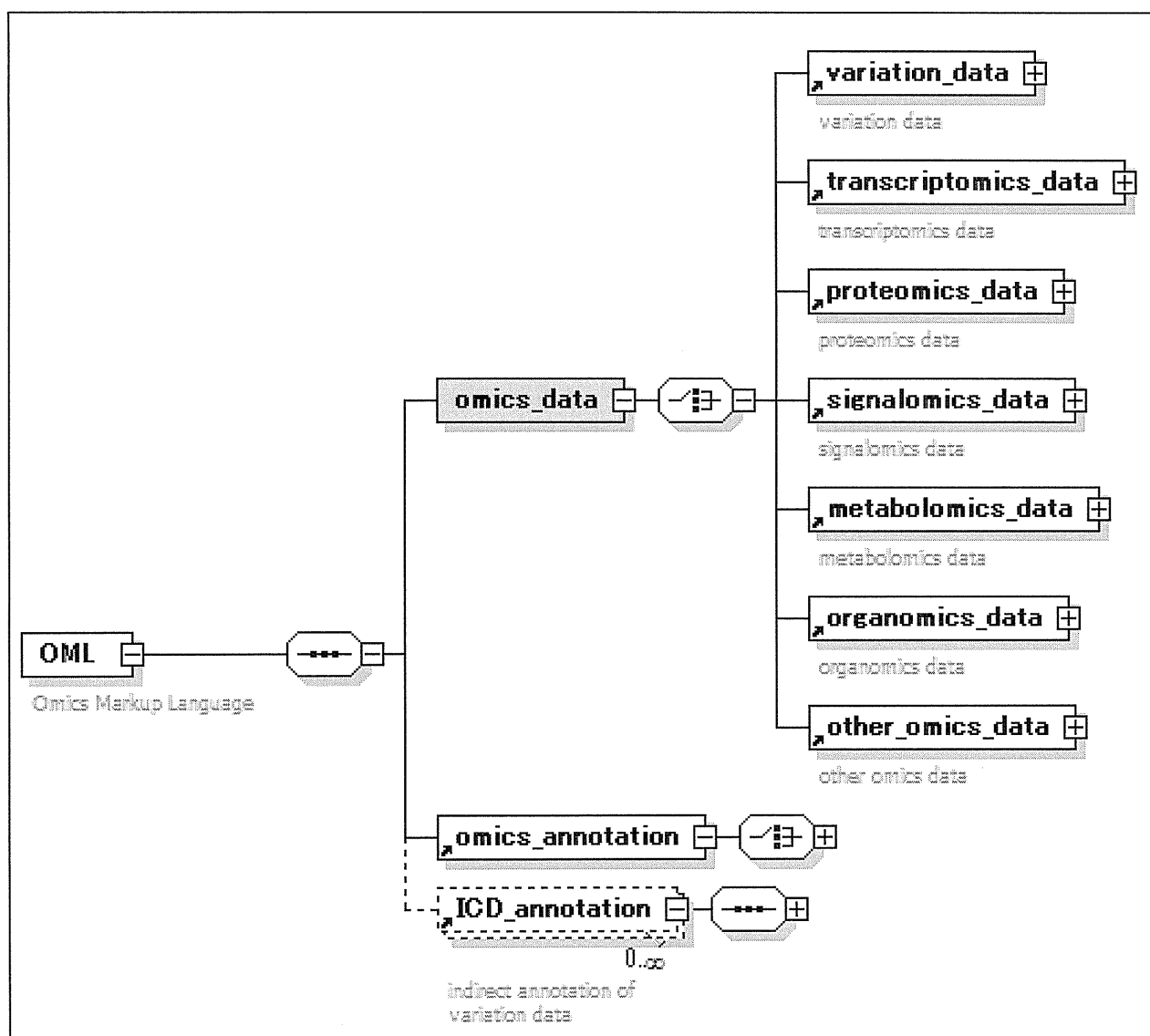
(図 I. ICD11_entity の概要 (第 1 階層まで展開))



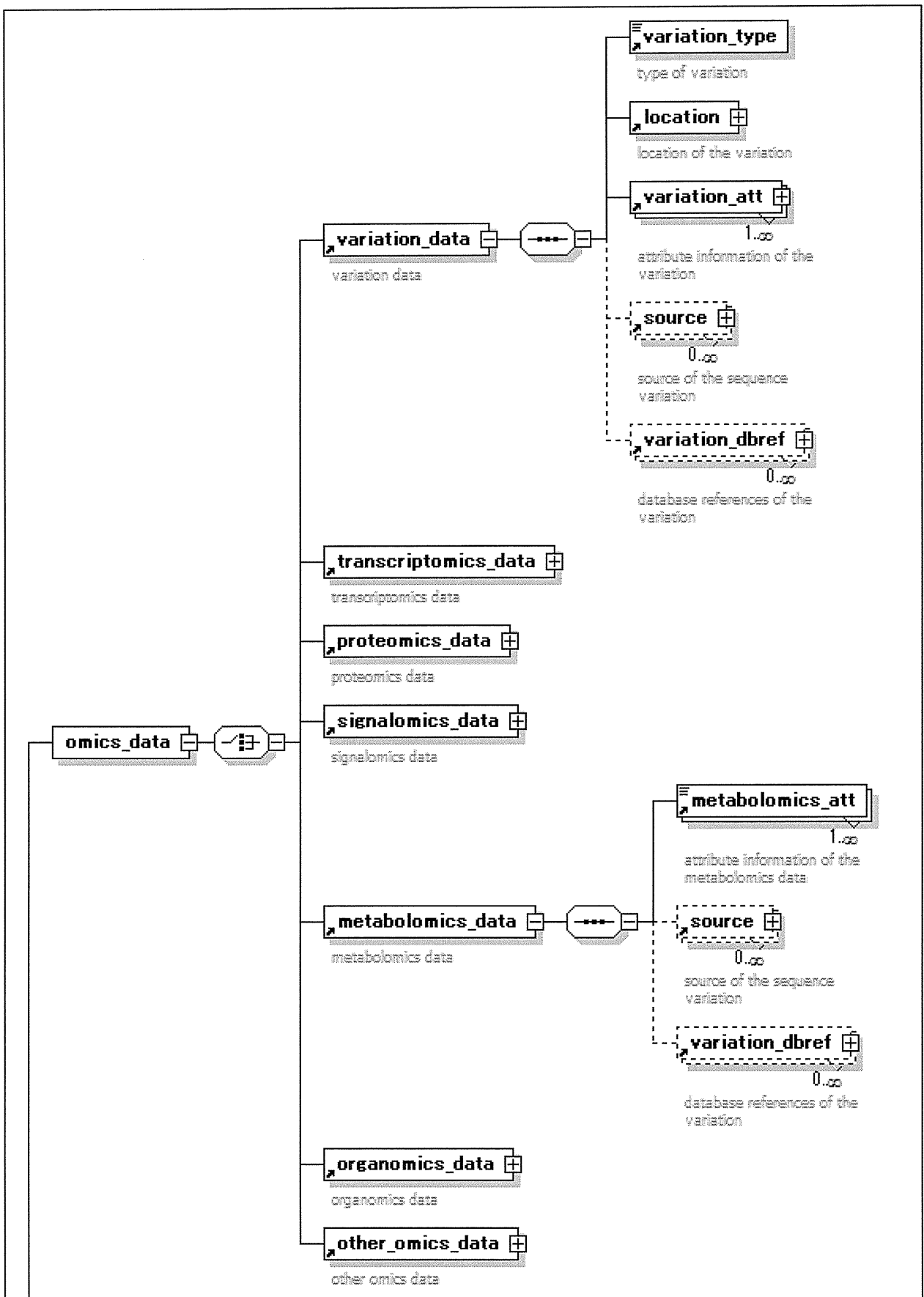
(図 J. Transcription_module (ICD11 ↔ OML 双方向転写モジュール) の概要)
(2 階層目まで展開)



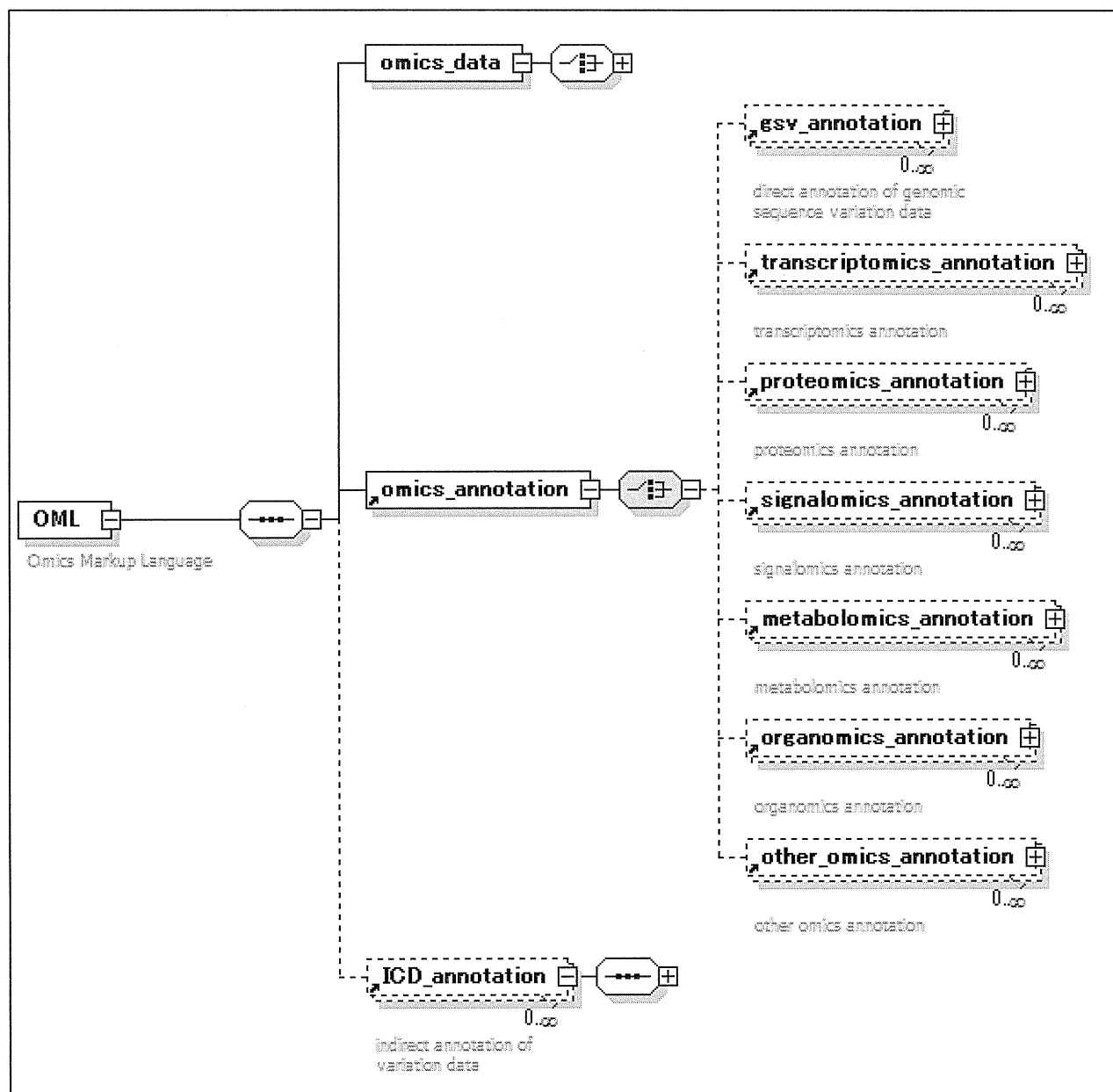
(図 K. OML_ICD11_Annotation_Group の構造)



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



(図 M. OML > Omics_data 以下の構造)



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