生命科学分野におけるオントロジーの活用事例: NANDO と PubCaseFinder

藤原 豊史
ライフサイエンス統合データベースセンター

自己紹介

藤原 豊史
ライフサイエンス統合データベースセンター
(DBCLS)

- 主に希少・遺伝性疾患に関するトランスレーショナルリサーチ
  - My works:
    - PubCaseFinder
      - DiseaseSearch
      - CaseSharing
      - PanelSearch
    - NANDO
    - NanbyoData
    - Mondo Japanese

本講義の概要

- 「オントロジーとは何か」という基本的な概念を解説し、生命科学分野におけるオントロジーの構築事例としてNANDOを、活用事例としてPubCaseFinderを紹介します。これらを通じて、オントロジーがデータ統合や検索の精度向上にどのように貢献するのかを具体的に説明します。

オントロジーの基本を学ぶために

- Gene Ontologyを構築したBarry Smithの講義資料が役に立つ
  - Barry Smithは、オントロジー（物ごとの名前や関係を整理する方法）分野における世界的に著名な研究者であり、医療や生命科学分野で実用的に活用されるバイオメディカルオントロジーの構築に貢献してきた
  - 講義資料
    - Introduction to Biomedical Ontologies
      - https://ontology.buffalo.edu/smith/BioOntology_Course.html

「Barry Smith Introduction to Biomedical Ontologies」で検索

オントロジーとは？

- 知識を整理するための「辞書＋ルール集」のようなもの
- バイオメディカルオントロジーは、生物・医療の用語や関係性をコンピュータで使える形でまとめたもの
- 例えば:
  - is_a: ～の一種（分類）
    - 「肺」は「臓器」の一つ → "肺 is_a 臓器"
  - part_of: ～の一部（構成）
    - 「心臓」は「循環器系の一部」→ "心臓 part_of 循環器系"

オントロジーとは？

- オントロジーで「意味」をたどる例
  - 「肺胞の異常」→「肺の異常」→「臓器の異常」
  - 「心室の異常」→「心臓の異常」→「循環器系の異常」

- part_of（構成）と is_a（分類）の関係を使って、症状がどのような構造・分類に属しているかを順にたどることができる

- 検索システムでの利用例：
  - 「肺胞の異常」と入力した場合、類似する「肺」や「臓器」レベルの情報も検索

⇨ 肺胞の異常から特定の肺疾患を候補として挙げたり、関連する遺伝子を提案することが可能になる

オントロジーとLLMはどう違う？どう補い合う？

- オントロジー：
  - 知識を明示的に定着させる（定義・分類・関係）
  - なぜそうなるかが明示的（is_a, part_of関係）
- LLM（大規模言語モデル）：
  - テキストから潜在的に探索し推論（定義されていなくても疾患・症状・治療などの関連性や、因果関係を推測できる柔軟さがある）
  - 出力の理由がブラックボックスになりやすい
- 補完関係
  - LLMが探索した知識や推論結果を、オントロジーによって整理・説明しやすい形に整えることができる

Kokash, Natallia, et al. "Ontology-and LLM-based Data Harmonization for Federated Learning in Healthcare." arXiv preprint arXiv:2505.20020 (2025).

実際に使われているオントロジー

- Gene Ontology（GO）
  - 生物の遺伝子産物の機能や関与するプロセスを体系的に整理
- Human Phenotype Ontology（HPO）
  - ヒトの疾患に関連する臨床症状を体系的に整理
- Monarch Disease Ontology（Mondo）
  - さまざまな疾患分類を統合し、疾患概念を体系的に整理
- Uberon
  - 多種にわたる動物の解剖学的構造を体系的に整理
- Nanbyo Disease Ontology（NANDO）
  - 指定難病・小児慢性特定疾病を病型分類まで体系的に整理

オントロジー構築事例：Nanbyo Disease Ontology

- 我々は指定難病制度および小児慢性特定疾病制度に関する難病およびその病型分類を体系的に整理した、**日本で初めての難病オントロジー** を構築
  - Nanbyo Disease Ontology（NANDO）

何のために難病オントロジーを構築したの？

- 国内の難病データと国外の疾患データとの相互運用性を高めるため

難病制度とは？

- 難病制度は、医療費助成などを通じて難治性疾患の患者を支援する枠組みであり、その中に**主に大人を対象**とした指定難病制度と、**子どもを対象**とした小児慢性特定疾病制度が位置づけられている

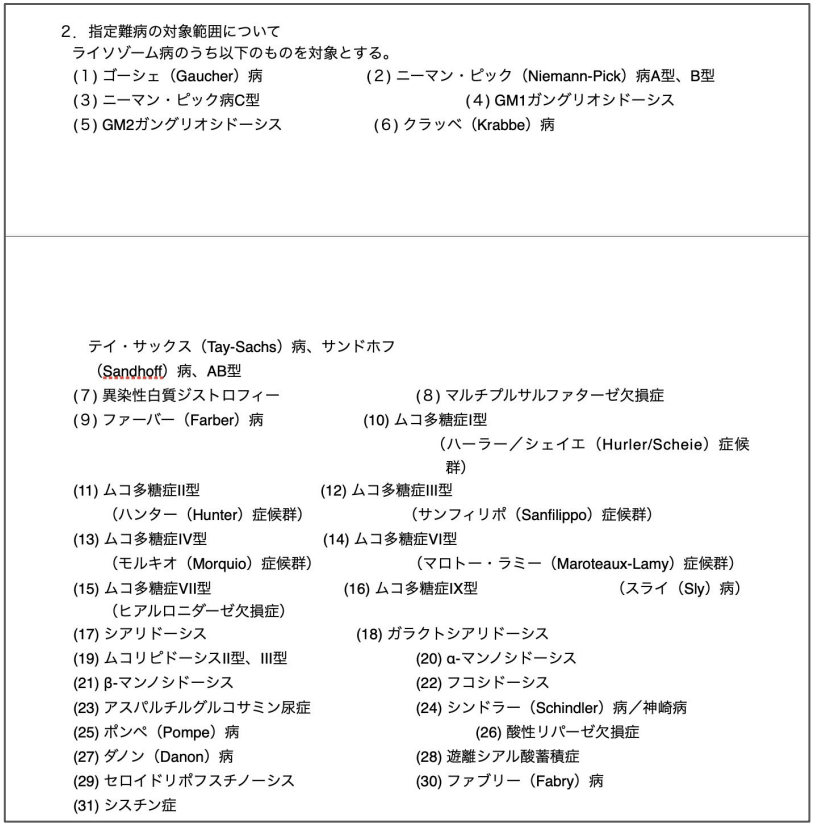
指定難病
- 難病法に基づいて定められた難治性疾患
- 原因不明で治療法が確立しておらず長期の療養が必要
- 発症頻度が低く、患者数が少ない
- 対象疾患は400種ほど
- 疾患ごとに認定基準が定められている

小児慢性特定疾病
- 児童福祉法に基づいて定められた難治性疾患
- 子どものうちに発症し、長期の治療や療養が必要な疾患
- 対象疾患は800種ほど
- 疾患ごとに認定基準が定められている

Nanbyo Disease Ontology（NANDO）

- 病型分類は公文書のなかに記載
  - ライソゾーム病：指定難病19
    - 病型分類として122疾患存在する

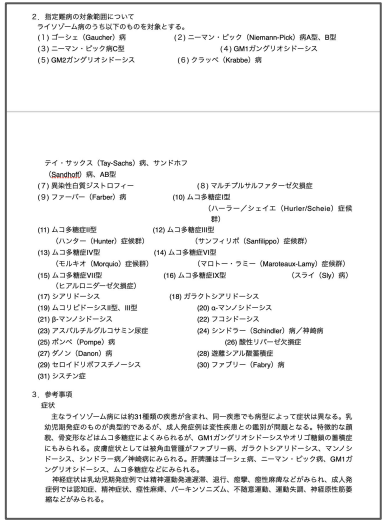
https://www.mhlw.go.jp/content/10905000/001173455.pdf



2．指定難病の対象範囲について
ライソゾーム病のうち以下のものを対象とする。
（1）ゴーシェ（Gaucher）病　（2）ニーマン・ピック（Niemann-Pick）病A型、B型
（3）ニーマン・ピック病C型　（4）GM1ガングリオシドーシス
（5）GM2ガングリオシドーシス　（6）クラッベ（Krabbe）病

テイ・サックス（Tay-Sachs）病、サンドホフ（Sandhoff）病、AB型
（7）異染性白質ジストロフィー　（8）マルチプルサルファターゼ欠損症
（9）ファーバー（Farber）病　（10）ムコ多糖症I型（ハーラー／シェイエ（Hurler/Scheie）症候群）
（11）ムコ多糖症II型（ハンター（Hunter）症候群）　（12）ムコ多糖症III型（サンフィリポ（Sanfilippo）症候群）
（13）ムコ多糖症IV型（モルキオ（Morquio）症候群）　（14）ムコ多糖症VI型（マロトー・ラミー（Maroteaux-Lamy）症候群）
（15）ムコ多糖症VII型（ヒアルロニダーゼ欠損症）　（16）ムコ多糖症IX型　（スライ（Sly）病）
（17）シアリドーシス　（18）ガラクトシアリドーシス
（19）ムコリピドーシスII型、III型　（20）α-マンノシドーシス
（21）β-マンノシドーシス　（22）フコシドーシス
（23）アスパルチルグルコサミン尿症　（24）シンドラー（Schindler）病／神崎病
（25）ポンペ（Pompe）病　（26）酸性リパーゼ欠損症
（27）ダノン（Danon）病　（28）遊離シアル酸蓄積症
（29）セロイドリポフスチノーシス　（30）ファブリー（Fabry）病
（31）シスチン症

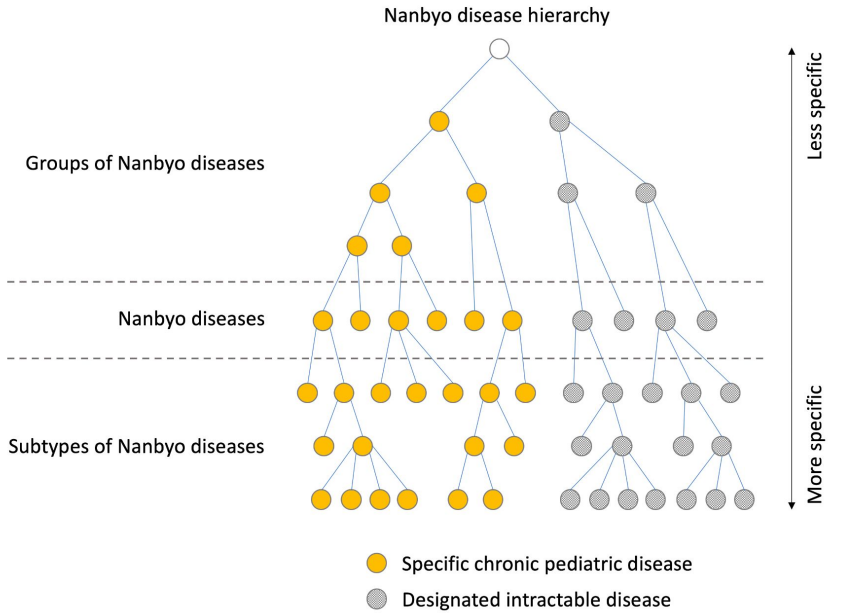
Nanbyo Disease Ontology（NANDO）



2．指定難病の対象範囲について
ライソゾーム病のうち以下のものを対象とする。
(1) ゴーシェ（Gaucher）病　(2) ニーマン・ピック（Niemann-Pick）病A型、B型
(3) ニーマン・ピック病C型　(4) GM1ガングリオシドーシス
(5) GM2ガングリオシドーシス　(6) クラッベ（Krabbe）病
テイ・サックス（Tay-Sachs）病、サンドホフ（Sandhoff）病、AB型
(7) 異染性白質ジストロフィー　(8) マルチプルサルファターゼ欠損症
(9) ファーバー（Farber）病　(10) ムコ多糖症I型（ハーラー／シェイエ（Hurler/Scheie）症候群）
(11) ムコ多糖症II型　(12) ムコ多糖症III型
（ハンター（Hunter）症候群）　（サンフィリポ（Sanfilippo）症候群）
(13) ムコ多糖症IV型　(14) ムコ多糖症VI型
（モルキオ（Morquio）症候群）　（マロトー・ラミー（Maroteaux-Lamy）症候群）
(15) ムコ多糖症VII型　(16) ムコ多糖症IX型　（スライ（Sly）病）
（ヒアルロニダーゼ欠損症）
(17) シアリドーシス　(18) ガラクトシアリドーシス
(19) ムコリピドーシスII型、III型　(20) α-マンノシドーシス
(21) β-マンノシドーシス　(22) フコシドーシス
(23) アスパルチルグルコサミン尿症　(24) シンドラー（Schindler）病／神崎病
(25) ポンペ（Pompe）病　(26) 酸性リパーゼ欠損症
(27) ダノン（Danon）病　(28) 遊離シアル酸蓄積症
(29) セロイドリポフスチノーシス　(30) ファブリー（Fabry）病
(31) シスチン症

3．参考事項
症状
主なライソゾーム病には約31種類の疾患が含まれ、同一疾患でも病型によって症状は異なる。乳幼児期発症のものが典型的であるが、成人発症例は変性疾患との鑑別が問題となる。特徴的な顔貌、骨変形などはムコ多糖症によくみられるが、GM1ガングリオシドーシスやオリゴ糖鎖の蓄積症にもみられる。皮膚症状としては被角血管腫がファブリー病、ガラクトシアリドーシス、マンノシドーシス、シンドラー病／神崎病にみられる。肝脾腫はゴーシェ病、ニーマン・ピック病、GM1ガングリオシドーシス、ムコ多糖症などにみられる。
神経症状は乳幼児期発症例では精神運動発達遅滞、退行、痙攣、個性喪失などがみられ、成人発症例では認知症、精神症状、痙性麻痺、パーキンソニズム、不随意運動、運動失調、神経原性筋萎縮などがみられる。

病型をis_aを用いて
体系的に整理



1,000件以上の公文書

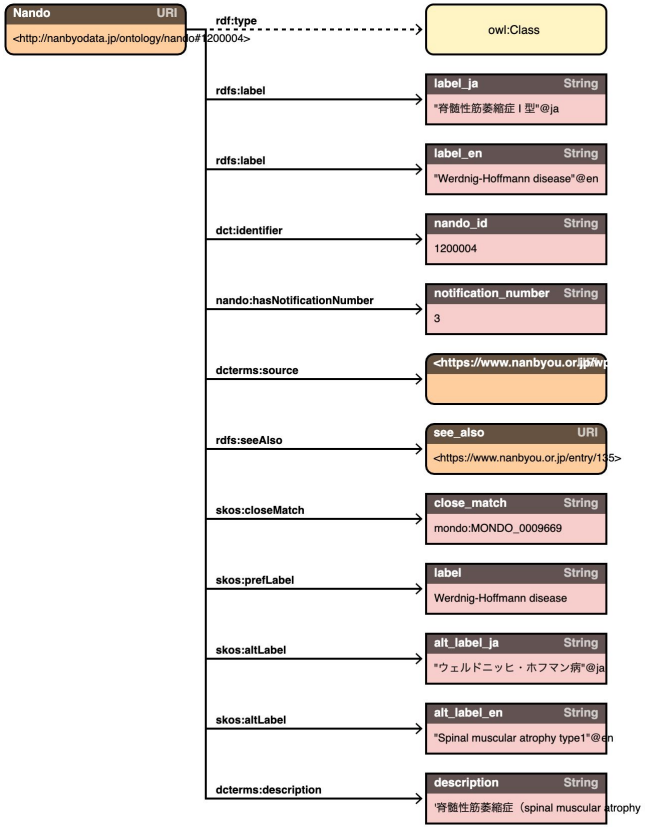
Nanbyo Disease Ontology（NANDO）

	指定難病	小児慢性特定疾病
# of groups of Nanbyo diseases	15	297
# of Nanbyo diseases	338	814
# of subtypes of Nanbyo diseases	746	565
Average number of children of Nanbyo diseases	2.2	0.69
Maximum number of children of Nanbyo diseases	110	20

Nanbyo Disease Ontology（NANDO）

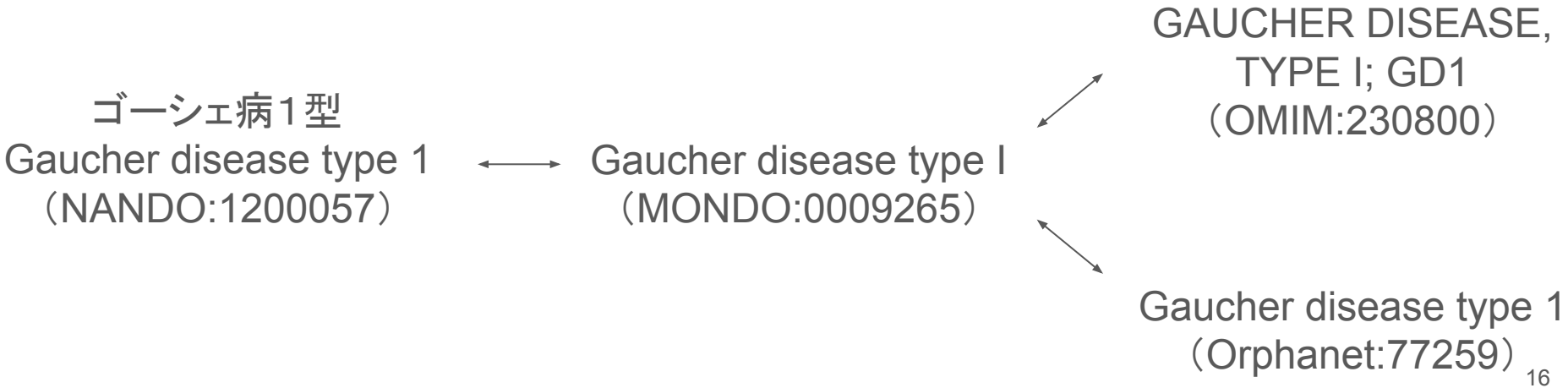
- それぞれの難病で以下を定義
  - ラベル（日・英）
  - シノニム
  - 疾患概要
  - 告示番号
  - 関連リソース

  など



NANDO↔Mondo マッピング

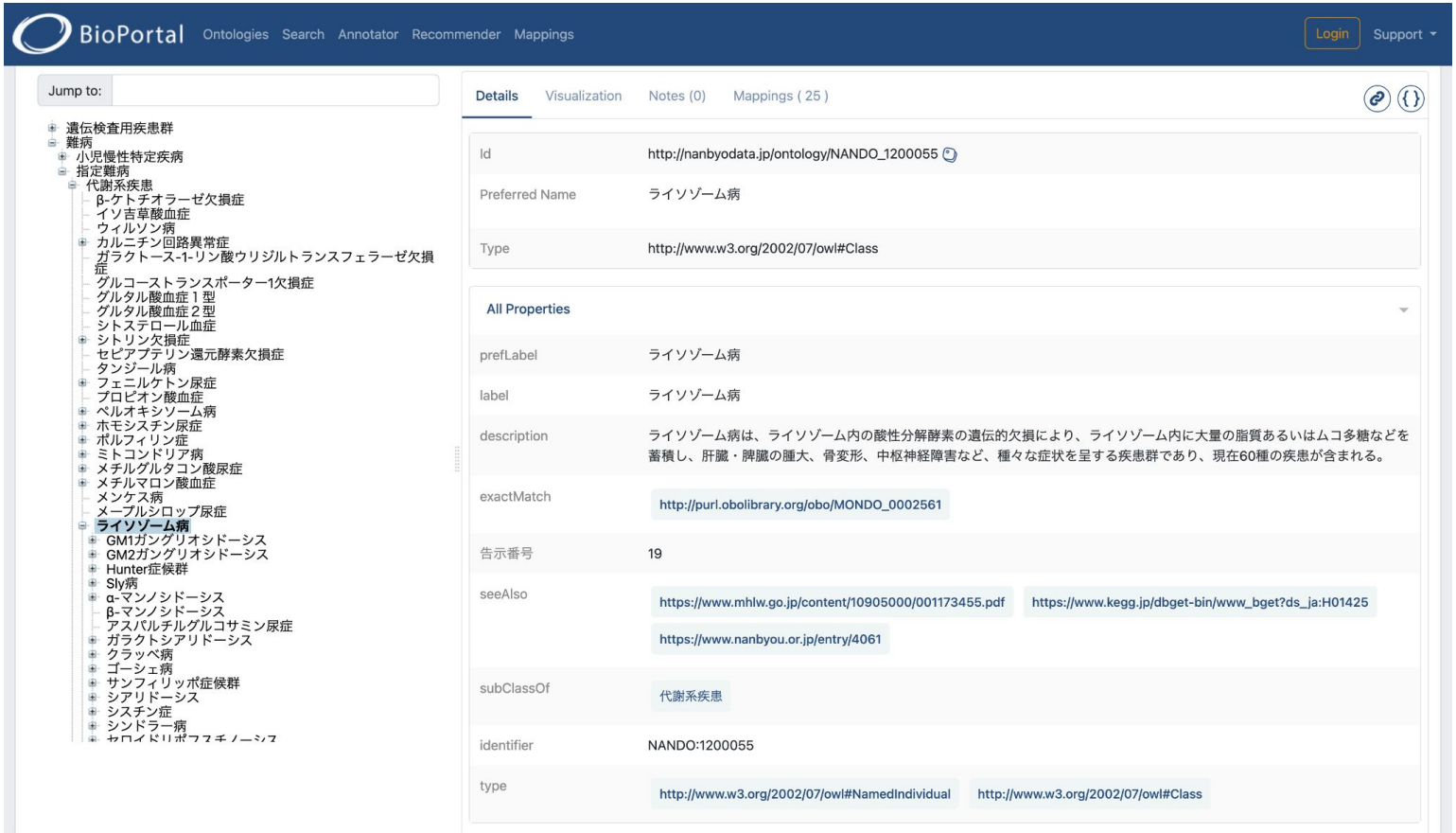
- 国内の難病データと国外の疾患データとの相互運用性を高めるために、NANDOを国際的な疾患オントロジーであるMondoにマッピング
  - 自動マッピング
  - 臨床医により自動マッピング結果を修正



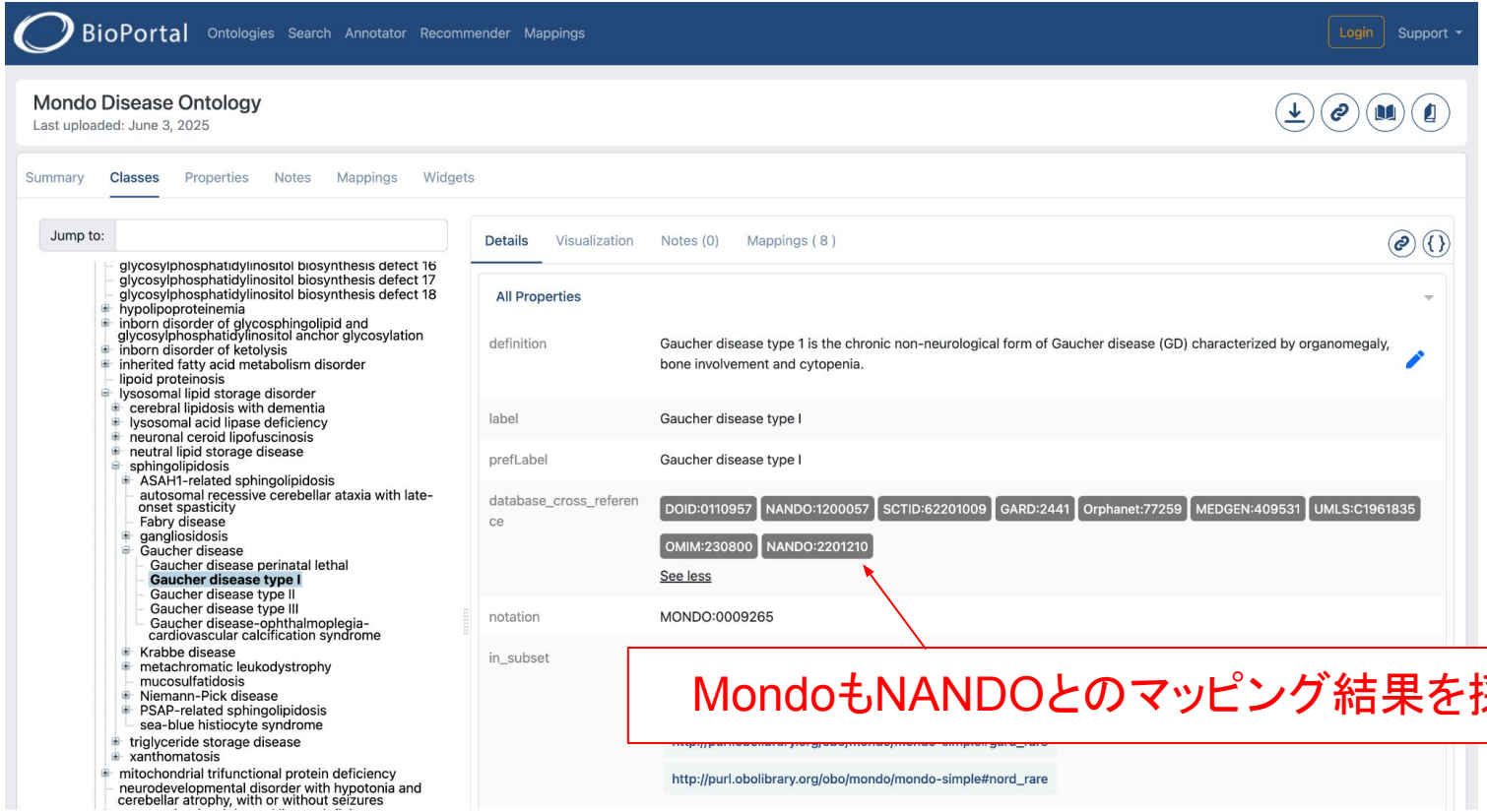
NANDO↔Mondo マッピング

	全疾患	指定難病	小児慢性特定疾病
# of terms	2,775	1,099	1,676
# of terms which include the cross-references to Mondo (Proportion of the total number of terms)	2,154 (77.6%)	892 (81.2%)	1,262 (75.3%)

Nanbyo Disease Ontology（NANDO）



MondoからNANDO（世界から日本）



Nanbyo Disease Ontology (NANDO)

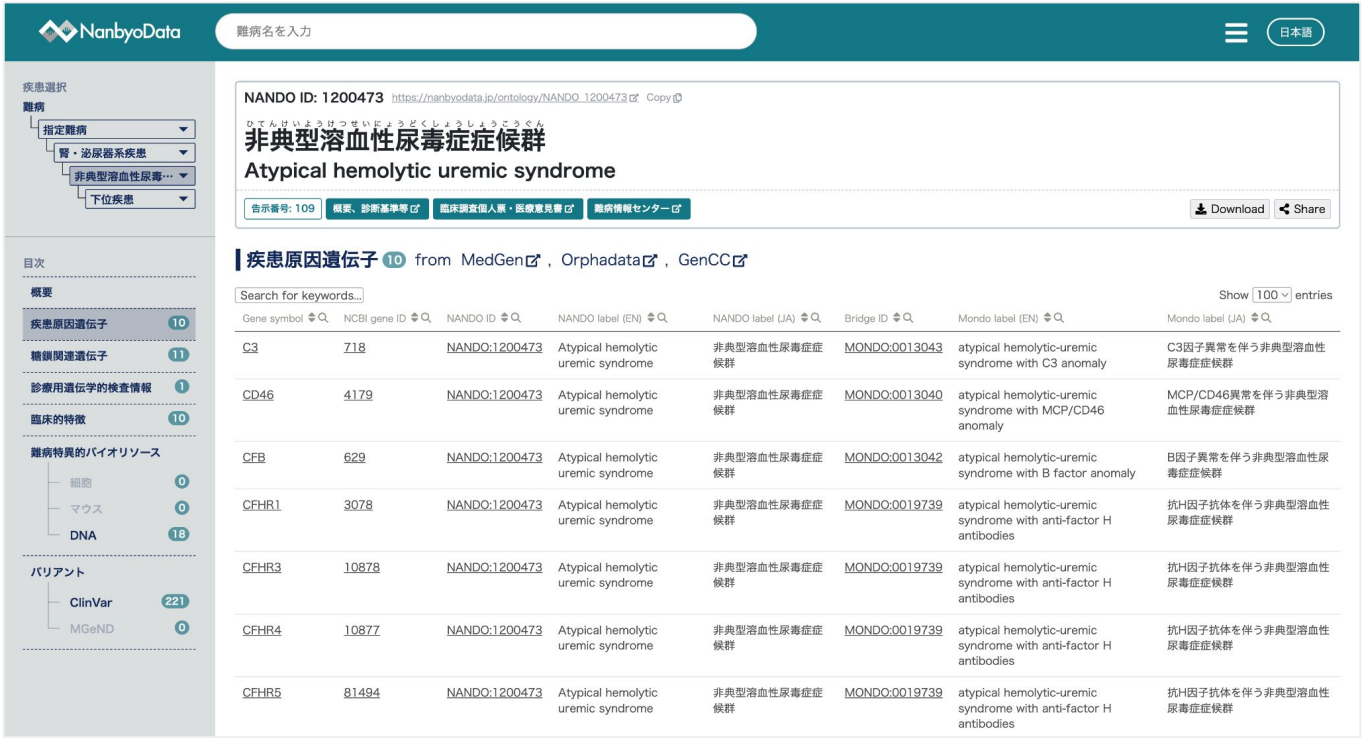
- https://nanbyodata.jp/about_nando

NanbyoData (https://nanbyodata.jp)

国内外のリソースから難病に関するデータを体系的に集約し、それらを閲覧・ダウンロード可能にするポータルサイト「NanbyoData」を開発

NanbyoData：国外データの例

疾患原因遺伝子をMedGen、Orphadata、GenCCから取得



NanbyoData：国内データの例

国内の診療用遺伝学的検査情報

日本人類遺伝学会
The Japan Society of Human Genetics

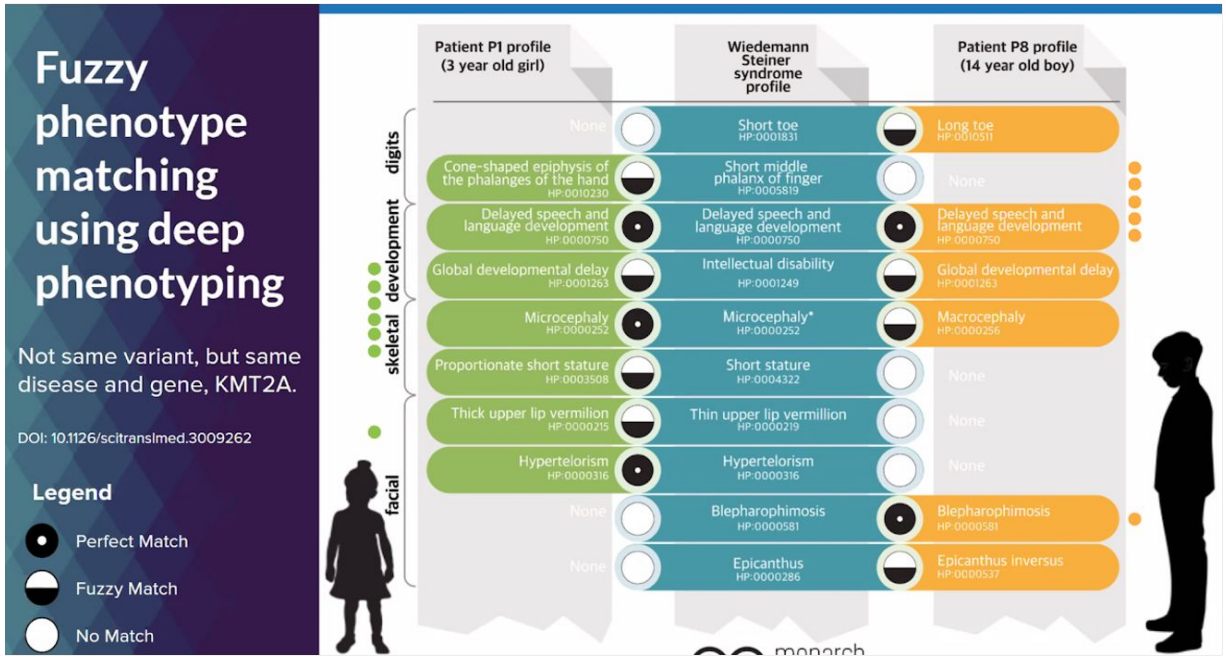
検査項目	非典型溶血性尿毒症症候群遺伝子検査
遺伝子	CFH,CFI,CD46,C3,CFB,THBD,DGKE
実施施設	かずさ遺伝子検査室
診療用	○
研究検査	
依頼方法	事前にメールで連絡 https://www.kazusa.or.jp/genetest/initial.html
保険点数	8,000
解析法	次世代シーケンサーあるいはキャピラリーシーケンサーによる遺伝子配列決定
検査に要する日数	検体到着後、 60 営業日以内
検体の採取量	血液 1 mL以上
検体の採取容器	EDTA-2K（または Na）真空密封型採血管

オントロジー構築事例：Nanbyo Disease Ontology

- NANDO/NanbyoDataを活用することで、国際的なデータ統合・共有を実現

オントロジー活用事例

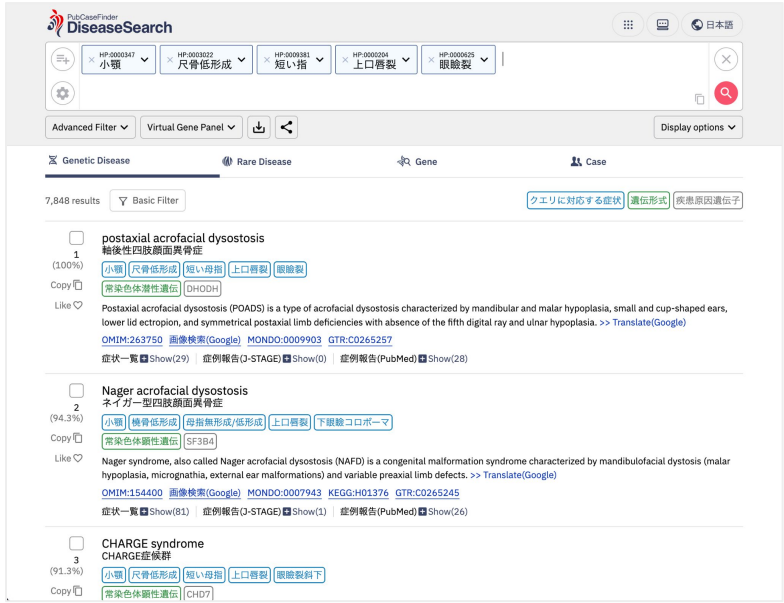
- 個々の希少・遺伝性疾患の患者数は少なく、疾患の種類数は1万件以上診療経験を積むことが難しく、診断が困難



Zemojtel, Tomasz, et al. "Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome." Science translational medicine 6.252 (2014): 252ra123-252ra123.

オントロジー活用事例：PubCaseFinder

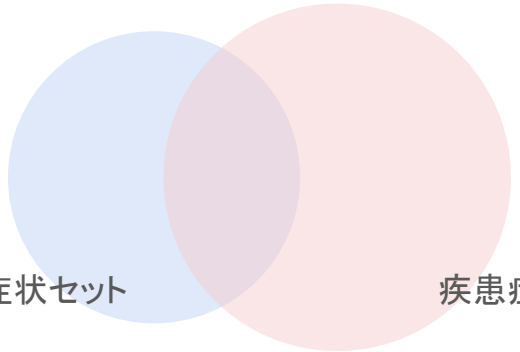
- 我々は希少・遺伝性疾患の診療支援を目的とした「PubCaseFinder」を開発



PubCaseFinder

- 臨床医・研究者が疾患ごとに関連症状をHPOで付与

類似度計算



入力症状セット　　疾患症状セット

postaxial acrofacial dysostosis
軸後性四肢顔面異骨症

小顎 | 尺骨低形成 | 短い母指 | 上口唇裂 | 眼瞼裂
常染色体潜性遺伝 | DHODH

Postaxial acrofacial dysostosis (POADS) is a type of acrofacial dysostosis characterized by mandibular and malar hypoplasia, small and cup-shaped ears, lower lid ectropion, and symmetrical postaxial limb deficiencies with absence of the fifth digital ray and ulnar hypoplasia. >> Translate(Google)

OMIM:263750 画像検索(Google) MONDO:0009903 GTR:C0265257

症状一覧 Hide(29) | 症例報告(J-STAGE) Show(0) | 症例報告(PubMed) Show(28)

乳房の異常

Add to Query	HPO ID	Label	Search
	HP:0002558	副甲状腺乳頭	

四肢の異常

Add to Query	HPO ID	Label	Search
	HP:0001374	先天性股関節脱臼	
	HP:0001159	合指趾症	
	HP:0003022	尺骨低形成	
	HP:0002974	橈尺骨癒合	
	HP:0002984	橈骨低形成	
	HP:0009778	短い母指	
	HP:0001760	足の異常	

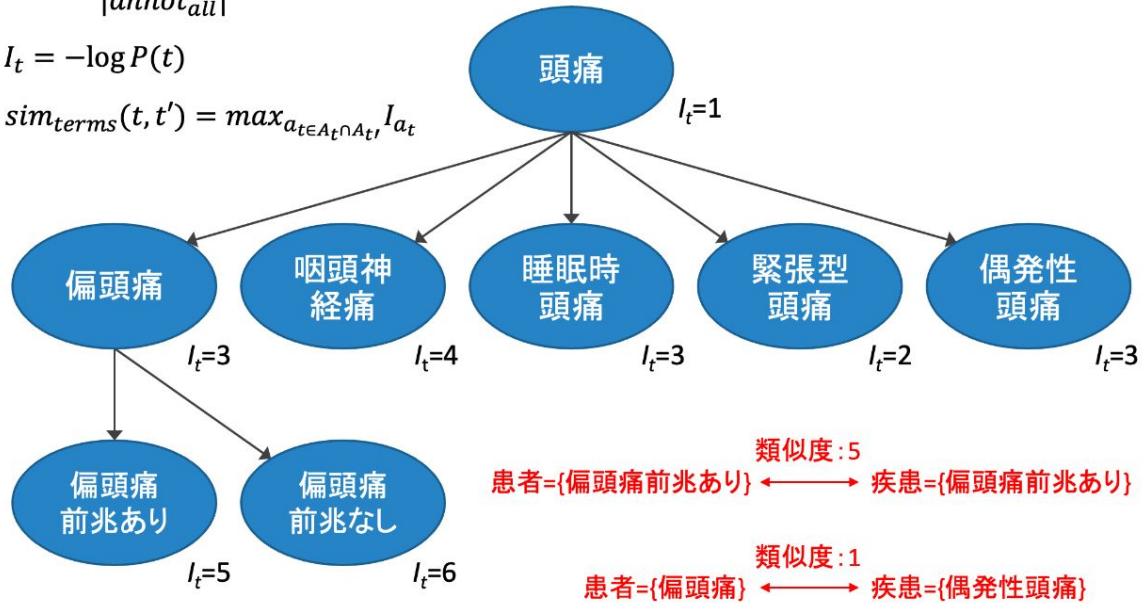
PubCaseFinder

- Human Phenotype Ontology（HPO）を用いて症状間の意味的な類似度を計算

$$P(t) = \frac{|annot_t|}{|annot_{all}|}$$

$$I_t = -\log P(t)$$

$$sim_{terms}(t, t') = max_{a_{t \in A_t \cap A_{t'}}} I_{a_t}$$



PubCaseFinder

- 患者の症状セットと疾患の症状セットとの類似度を計算

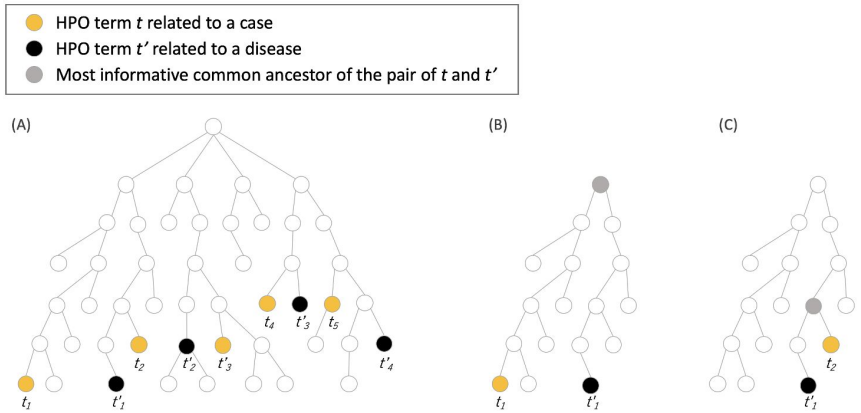
$$P(t) = \frac{|annot_t + 1|}{|annot_{all} + 1|}, IC_t = -\log P(t)$$

$$sim(t, t') = max_{a_t \in A_t \cap A_{t'}} IC_{a_t}$$

$$SIM(c, d) = \frac{\sum_{t \in c} R_t \times max_{t' \in d} sim(t, t')}{\sum_{t \in c} R_t \times IC_t}$$

$$C(c, d) = \{t \in c | \Delta t < E\}$$

$$R_t = \begin{cases} 1 & |C(c, d)| < K \ or \ \Delta t < E \\ W & otherwise \end{cases}$$

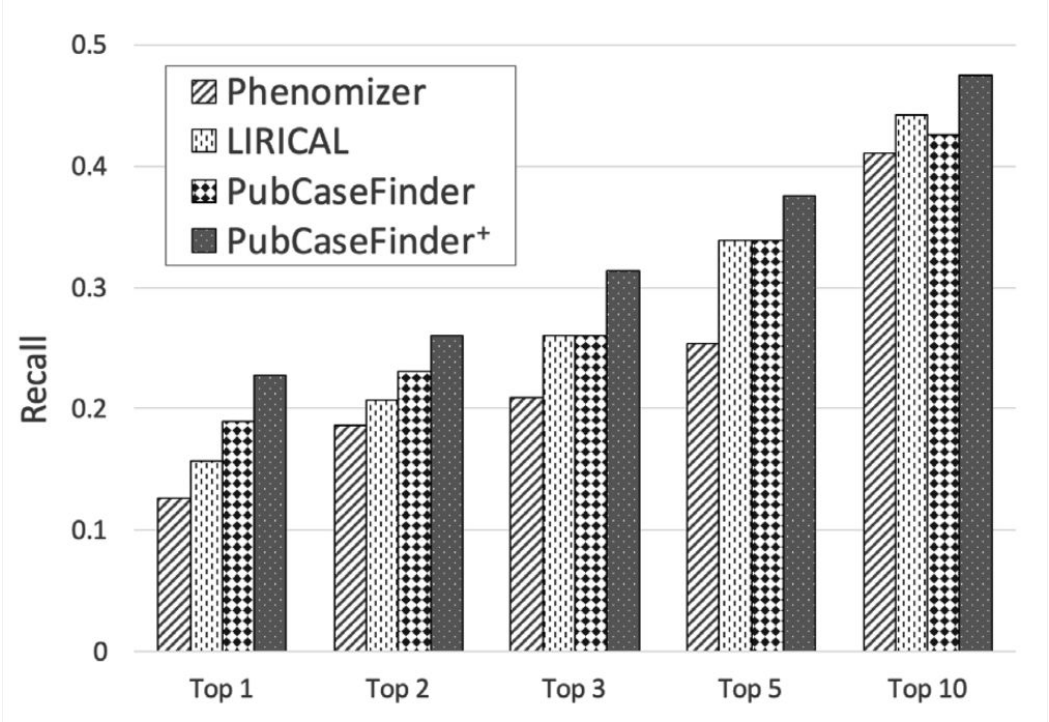


Fujiwara, T., Shin, J. M., & Yamaguchi, A. (2022). Advances in the development of PubCaseFinder, including the new application programming interface and matching algorithm. Human mutation, 43(6), 734–742. https://doi.org/10.1002/humu.24341

PubCaseFinder

- 疾患ランキング精度

藤原豊史; 仁宮洸太. オントロジーを用いた希少・遺伝性疾患診断支援人工知能, (2020), 35.4: 480-486.

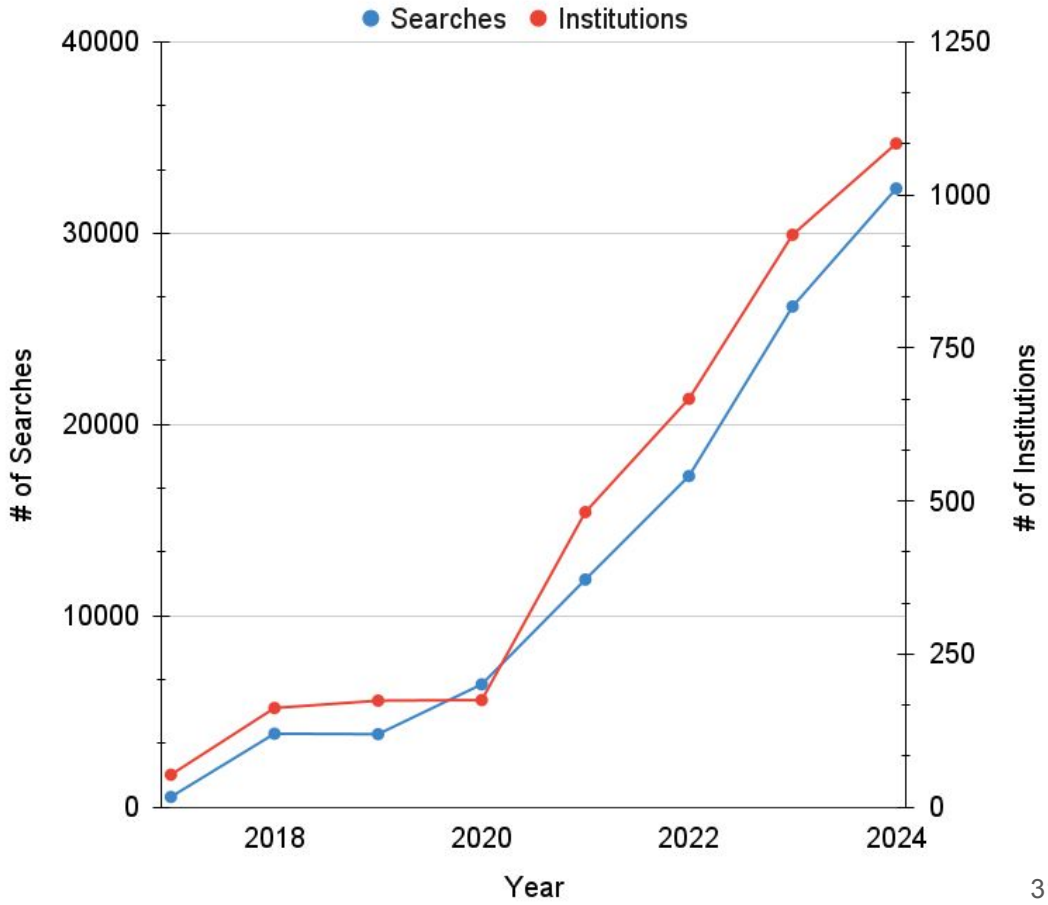


PubCaseFinder

- アクセス数
  - 250万ページビュー（2024年度）

【対象施設】
施設名に以下のwordを含む

university
college
hospital
medical
clinical center
iryou
dental
treatment center
cancer center
pharma
gakuen
school
laboratory
national
ministry
dokuritsu gyousei



まとめ

- 国内の難病データと国外の疾患データとの相互運用性を高めるための難病オントロジー「NANDO」を紹介
  - 指定難病制度および小児慢性特定疾病制度に関する難病およびその病型分類を日本で初めて体系的に整理したオントロジー
  - NANDOを用いて難病ポータルサイト「NanbyoData」を構築し、難病データの国際的な統合・共有を実現

- 希少・遺伝性疾患の診療支援を目的とした「PubCaseFinder」を紹介
  - HPO、Mondoなどのオントロジーを活用したウェブサービスで、世界の多くの臨床医が活用