

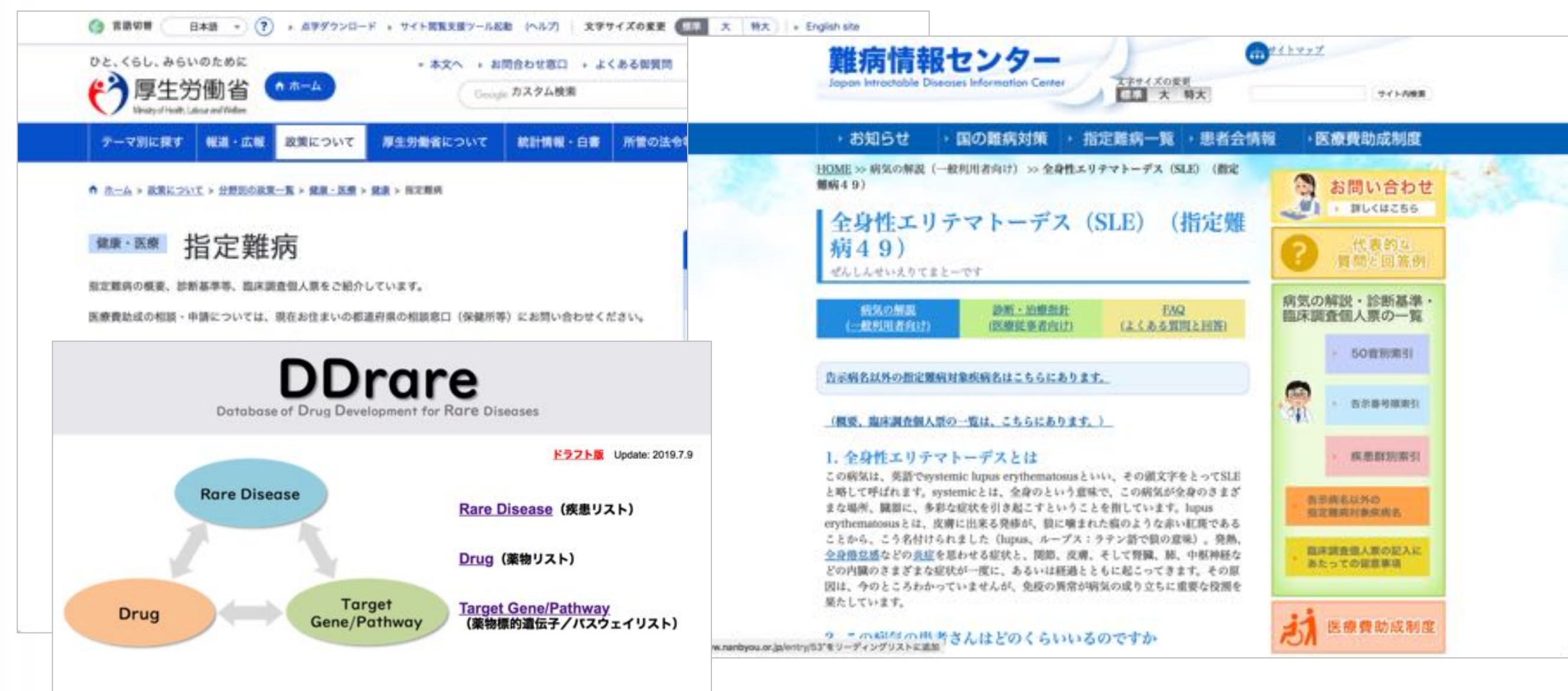
難病・希少疾患の疾患オントロジー

○平田誠¹⁾ 坂手龍一¹⁾ 深川明子²⁾ 木村友則¹⁾

¹⁾ 国立研究開発法人 医薬基盤・健康・栄養研究所 難治性疾患研究開発・支援センター ²⁾ 同研究所 AI健康・医薬研究センター

難病・希少疾患の多くは患者数が少ないため疾患の定義が難しく、国内外のデータベース (DB) 間で疾患名表記の違いが見られる。本研究では指定難病 (2019年7月の333疾患) の英語名リストを作成し、DB間 (指定難病 (日)、GARD (米)、Orphanet (欧) が主な対象) での疾患名の対応付けを試みた。難病・希少疾患の疾患名についても、別名、派生名、省略形を含む、階層構造を持つ情報体系としての整理、そして、疾患定義の医学的な見地でのアノテーションが、より重要になるだろう。

指定難病疾患名のDB間での対応付け



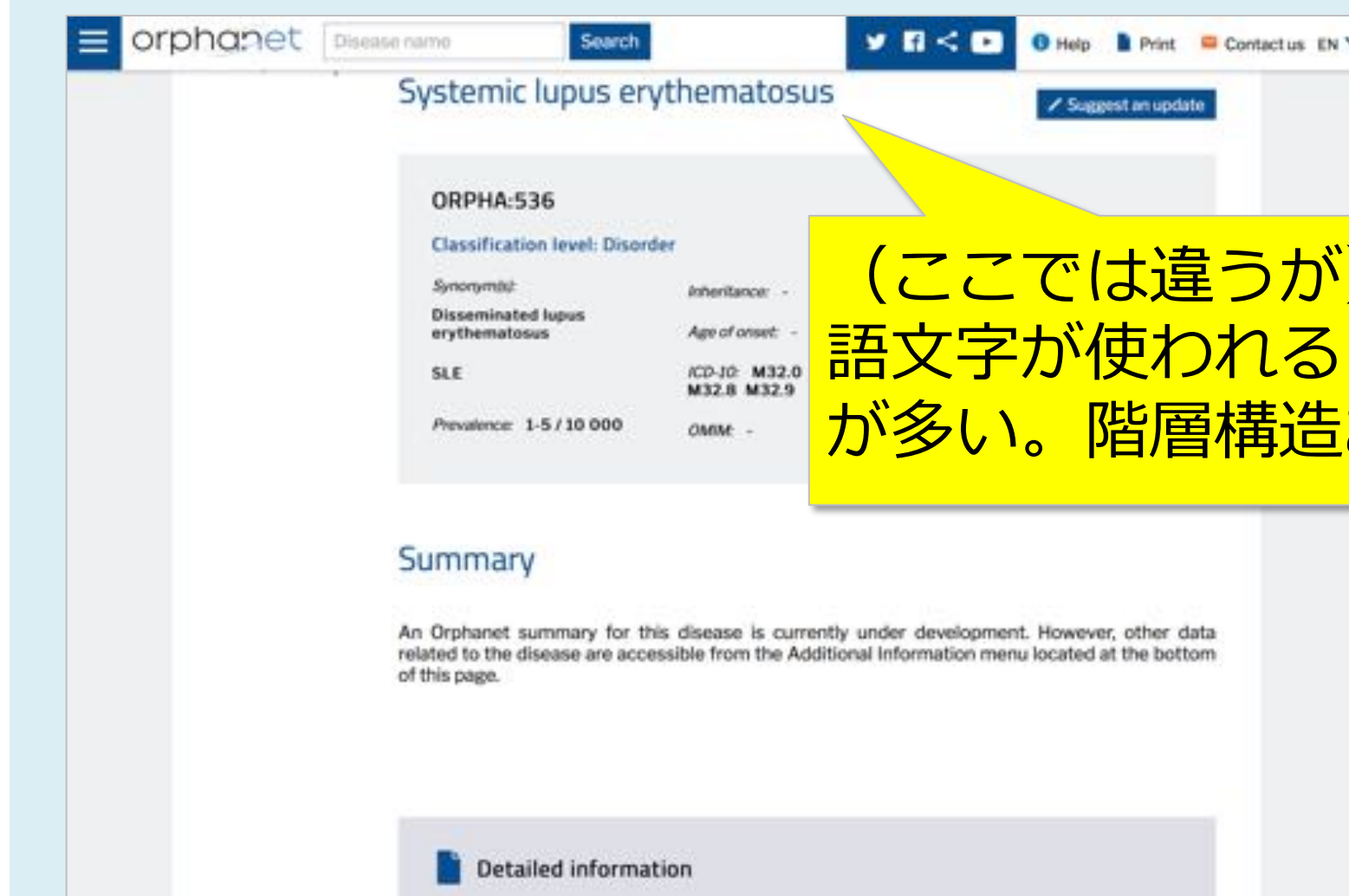
指定難病

情報は厚生労働省のサイトや難病情報センターから取得可能。医薬基盤・健康・栄養研究所では臨床試験や薬物、薬物標的遺伝子・パスウェイ情報をDDrare (データベース: <https://ddrare.nibiohn.go.jp>) から公開。

解析方法

- 指定難病の英語名: DDrare (現在331疾患) を利用 (2疾患を別途追加して計333疾患)
- 各DBの疾患の代表名、同義語などを対象に、完全一致相当で比較
- 指定難病と小児慢性特定疾病との対応: 世田谷区ウェブサイト (<https://www.city.setagaya.lg.jp>) より取得 (2019年2月時点)
- 文字処理: 「s」や「-」などを削除, 大文字・小文字無視, 仏独文字などを英字変換 (例: Behçet disease -> Behcet disease) など

各DBのデータ 例) 全身性エリテマトーデス (Systemic lupus erythematosus)



Orphanet
<https://www.orpha.net>



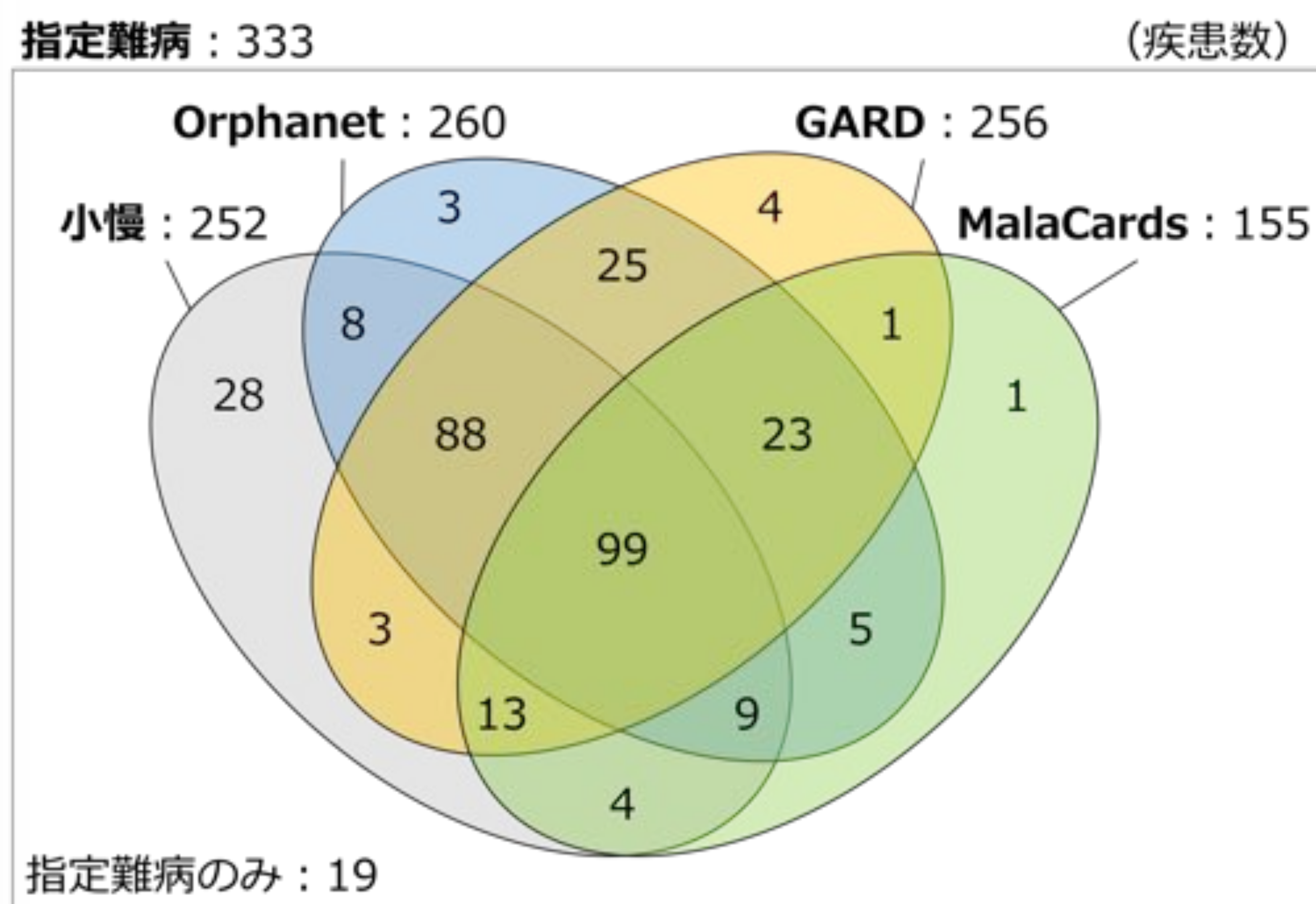
Genetic and Rare Diseases Information Center (GARD)
<https://rarediseases.info.nih.gov>



小児慢性特定疾病情報センター
<https://www.shuman.jp>



MalaCards: The human disease database
<https://www.malacards.org>



指定難病の4DBとの対応

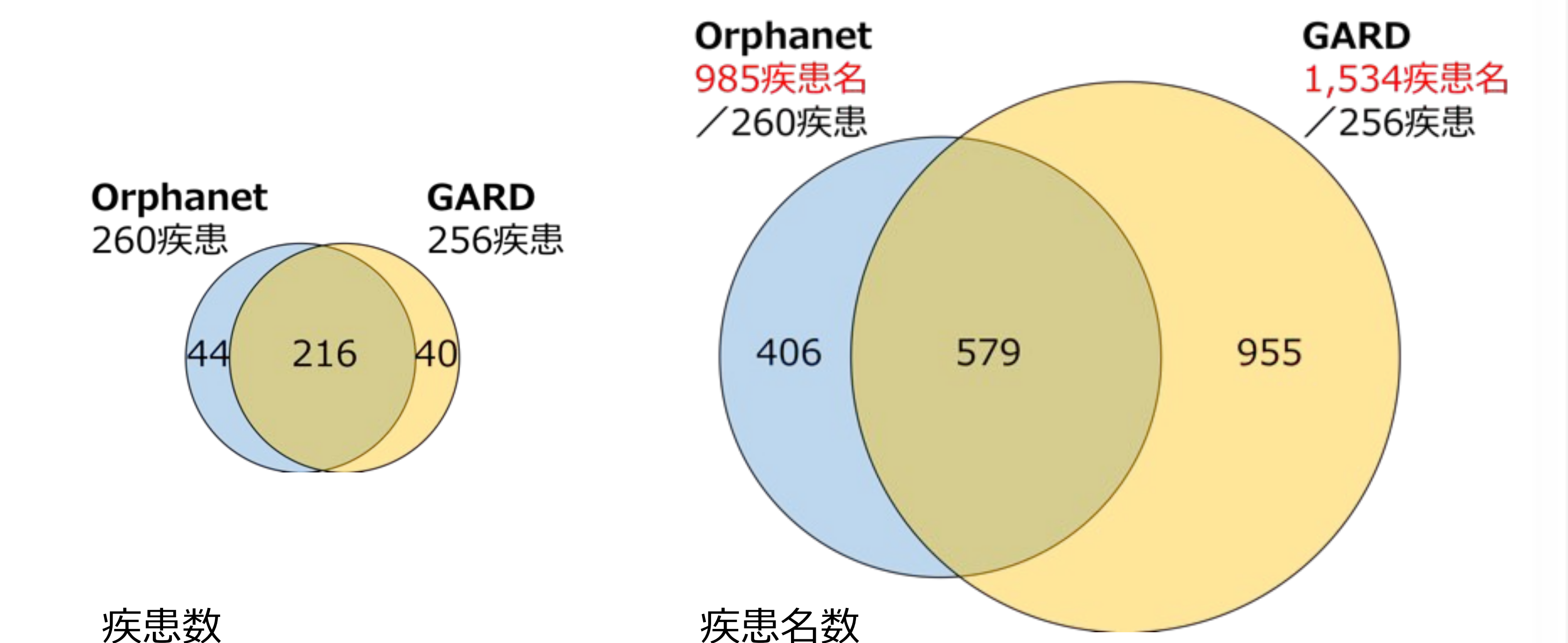
難病・希少疾患に特化した日米欧3つのDB (MalaCardsを除く) に対して、指定難病333疾患の75%前後について対応付けできた。OrphanetとGARDは互いに90%以上の疾患が対応付くのにに対し、指定難病と小児慢性特定疾病 (小慢) では76%にとどまるが、これは後者が小児疾患に特化していることも原因である (例: 主として成人に発症するハンチントン病がない)。

考察

- 難病・希少疾患領域は、新病名の創出 (診断基準の確率) や疾患名のアップデートが比較的頻繁な傾向があり、目的に応じたDB・辞書の利用が推奨される。
- 疾患名表記の揺れ (「~を除く」など自由表記に近いものなど)、階層構造 (Type 1a, Familial/Sporadicなど)、省略語の非特異性 (SMA, MSなど) など、対応付けの技術的課題については、テキストマイニングにおける機械学習の活用が期待される。
- 情報解析技術だけでなく、疾患の対応付けの根拠として、医学・生物学的な知見 (疾患発症機序、診断基準、日本人の特異性など) が重要である。
- 指定難病については、まずは小児慢性特定疾病との対応と英語名の定義を期待したい。

その他の疾患オントロジー関連DB・辞書 (たくさんある!)

Disease Ontology, Experimental Factor Ontology, Human Disease Ontology, ICD9/10/11, ICD9CM, KEGG, LSD, Mondo Disease Ontology, MeSH, MedDRA, MedGen, NCI, NIFSTD, OMIM, OMIMPS, ORDO, SNOMED-CT, UMLS, 万病辞書, 標準病名マスター, ...



OrphanetとGARDの比較 (別名、派生名、省略形などの数)

OrphanetとGARDは、指定難病と対応する疾患数はほぼ同じだが、それぞれ1疾患あたり別名などの複数の疾患名がある。1疾患あたりの平均疾患名数は、Orphanet: 3.8, GARD: 6.0。GARDは各疾患の別名などの数が多い、Orphanetの1.5倍以上。

指定難病の省略語について

文献にもとづく省略語DBのAllie (<https://allie.dbcls.jp>) からは、指定難病301疾患について、計959の省略語 (1疾患あたり平均3.2語) が得られた (9月25日時点)。