



Analyzing rare diseases terms in biomedical terminologies

Lister Hill National Center for Biomedical Communications

Erika Pasceri, Bastien Rance, Olivier Bodenreider

December 17, 2010

What is a rare disease?

A *rare disease* is a pathological condition with low prevalence and incidence.

There are between 6000 and 8000 rare diseases.

Many rare diseases are sparsely distributed in *some geographic areas* and *more frequent in others*, for reasons linked to genetic factors, environmental conditions that influence the spread of pathogens and the life habits.

Thalassemia, for example, is a relatively common genetic disease in the Mediterranean basin (very common in Southern Italy) and rare in the United States.

Rare diseases portals: Office of Rare Diseases Research

The screenshot shows the GARD website interface. At the top is a navigation menu with categories: Rare Diseases Information, Patient Advocacy Groups, Research & Clinical Trials, Genetic & Rare Diseases Information Center (highlighted), and Scientific Conferences. Below this is a secondary menu with: Genetics Information & Services, Research Resources, Patient Travel & Lodging, Reports & Publications, Rare Diseases News, and Recursos en español.

Genetic and Rare Diseases Information Center (GARD)

- GARD Home
- About GARD
- Contact GARD
- Feedback for GARD
- ORDR Home
- NHGRI Health Information Page

Home > Genetic & Rare Diseases Information Center (GARD) > Cryptococcosis

Cryptococcosis

Please note that the links contained on this search results page may take you to sites outside of the NIH. (See [Disclaimer](#) under Site Policies for details.)

These Web pages are updated as the Genetic and Rare Diseases Information Center receives questions and as new information becomes available. If you don't see many information resources on this page, it may be because the Information Center hasn't yet received a question about this condition.

If you have a question, please [contact us](#) — we will answer your question and update this page with new resources and information.



Questions & Answers

If you would like to submit a question, [Contact GARD](#)

[Show All Resources](#)

For more information about Cryptococcosis click on the boxes below:

NLM Gateway	Medical Products	Scientific Conferences	Support Groups	Clinical Trials & Research
--------------------	-------------------------	-------------------------------	-----------------------	---------------------------------------

NLM Gateway (Found 1 resource)

A tool to search across multiple resources offered on the National Library of Medicine's Website

- [NLM Gateway](#) allows users to search many resources offered on the National Library of Medicine's Web site at once to quickly find more information about this condition. Some of the resources may be a duplicate of the resources listed on this page. To search NLM Gateway, click on the link; the condition name will already be in the search box, so you can just click the "Search" button.

ORDR (Office of Rare Diseases Research)

<http://rarediseases.info.nih.gov/Default.aspx>

Rare diseases portals: Orphanet

The screenshot shows the Orphanet website interface. At the top, there are language options: Français, English (selected), Español, Deutsch, Italiano. The Orphanet logo is on the left, and navigation links (Homepage, Help, Contact us) are on the right. A horizontal menu contains categories: Rare diseases, Orphan drugs, Expert centres, Diagnostic tests, Research and trials, Patient organisations, Directory of resources, and Other information. Below this is a secondary menu with search options: Search, Search by sign, Classifications, Genes, Encyclopaedia for patients, Encyclopaedia for professionals, and Emergency guidelines. The main content area shows a search path: Homepage » Rare diseases » Search. A search box contains 'cryptococcosis' and is marked as a mandatory field. Search criteria are set to 'Disease name'. An 'OTHER SEARCH OPTION(S)' section offers an 'Alphabetical list'. The search results for 'Cryptococcosis' are displayed in a table-like format:

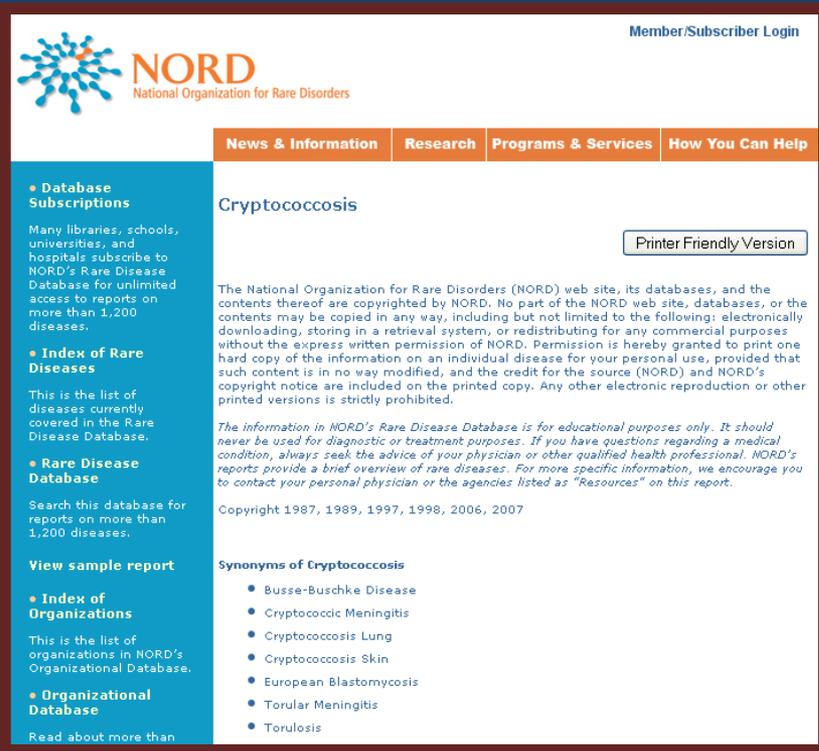
Orpha number	: ORPHA1546	Synonym(s)	: -
Prevalence of rare diseases	: Unknown		
Inheritance	: Sporadic		
Age of onset	: Variable		
ICD 10 code	: B45		
MIM number	: -		

Below the table is a 'SUMMARY' section with a text description of Cryptococcosis. To the right of the summary is an 'Additional information' section with a sub-section 'Further information on this disease' containing a list of links: Classification(s) (1), Gene(s) (0), and Other website(s) (4).

Orphanet

<http://www.orpha.net/consor/cgi-bin/index.php>

Rare diseases portals: National Organization of Rare Disorders



Member/Subscriber Login

NORD
National Organization for Rare Disorders

News & Information | Research | Programs & Services | How You Can Help

• **Database Subscriptions**
Many libraries, schools, universities, and hospitals subscribe to NORD's Rare Disease Database for unlimited access to reports on more than 1,200 diseases.

• **Index of Rare Diseases**
This is the list of diseases currently covered in the Rare Disease Database.

• **Rare Disease Database**
Search this database for reports on more than 1,200 diseases.

View sample report

• **Index of Organizations**
This is the list of organizations in NORD's Organizational Database.

• **Organizational Database**
Read about more than

Cryptococcosis

[Printer Friendly Version](#)

The National Organization for Rare Disorders (NORD) web site, its databases, and the contents thereof are copyrighted by NORD. No part of the NORD web site, databases, or the contents may be copied in any way, including but not limited to the following: electronically downloading, storing in a retrieval system, or redistributing for any commercial purposes without the express written permission of NORD. Permission is hereby granted to print one hard copy of the information on an individual disease for your personal use, provided that such content is in no way modified, and the credit for the source (NORD) and NORD's copyright notice are included on the printed copy. Any other electronic reproduction or other printed versions is strictly prohibited.

The information in NORD's Rare Disease Database is for educational purposes only. It should never be used for diagnostic or treatment purposes. If you have questions regarding a medical condition, always seek the advice of your physician or other qualified health professional. NORD's reports provide a brief overview of rare diseases. For more specific information, we encourage you to contact your personal physician or the agencies listed as "Resources" on this report.

Copyright 1987, 1989, 1997, 1998, 2006, 2007

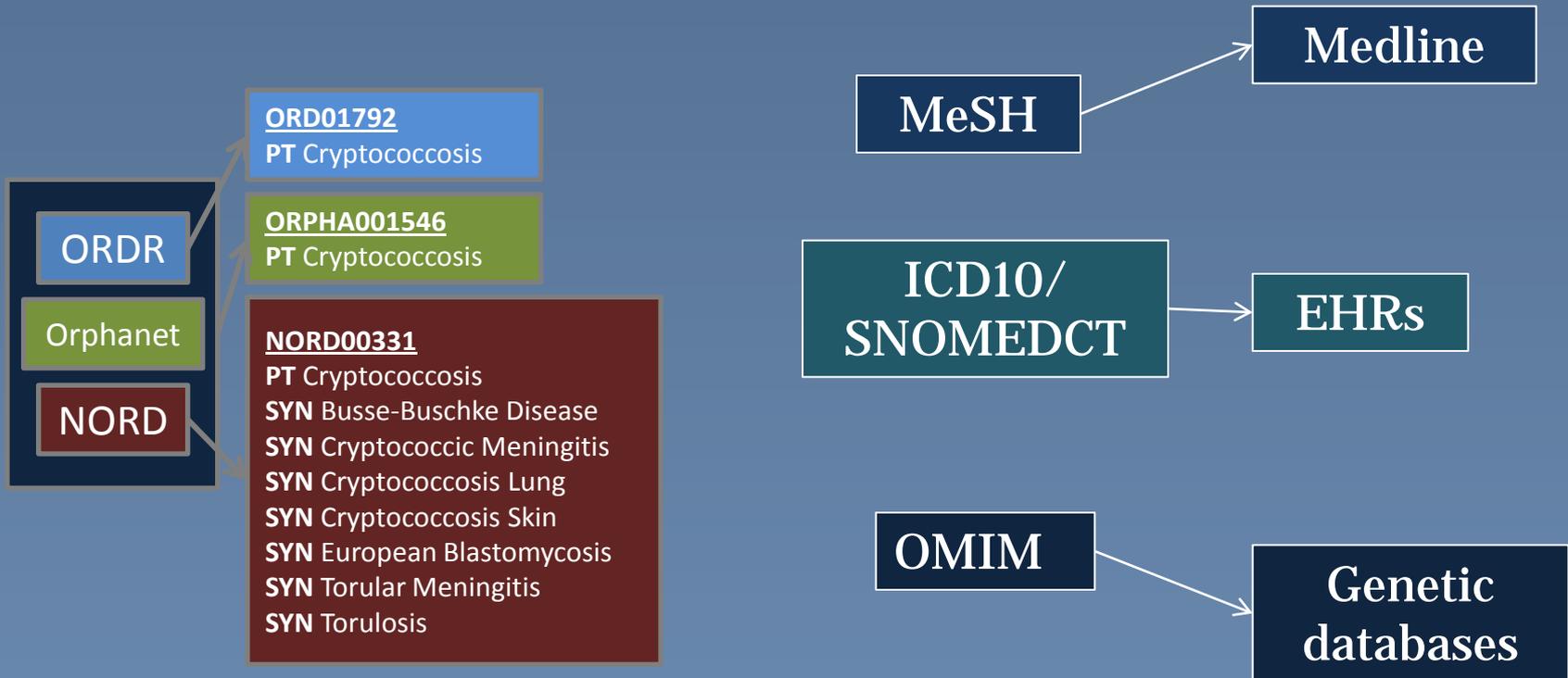
Synonyms of Cryptococcosis

- Busse-Buschke Disease
- Cryptococcic Meningitis
- Cryptococcosis Lung
- Cryptococcosis Skin
- European Blastomycosis
- Torular Meningitis
- Torulosis

NORD (National Organization for Rare Disorders)

<http://www.rarediseases.org/>

Overview



Objectives

Examine representation of rare diseases terms in biomedical terminologies (through UMLS), such as:

- **ICD10** (International Classification of Diseases 10th revision)
- **MeSH** (Medical Subject Headings)
- **OMIM** (Online Mendelian in Man)
- **SNOMED CT** (Systematized Nomenclature of Medicine-Clinical Terms)

Objectives (2)

- Overlap among sources
- Presence / absence in target sources in the Unified Medical Language System (UMLS):
 - At the concept level
 - At the term level
- For a given disease acquire more specific information:
 - Synonyms
 - Descendants

Sources

▪ ORDR

Tot. Concepts = 6,857

Tot PT = 6,857

Tot. Syn. = 11,803

A rare (or orphan) disease is generally considered to **have a prevalence of fewer than 200,000 affected individuals in the United States.** Certain diseases with 200,000 or more affected individuals may be included in this list if certain subpopulations of people who have the disease are equal to the prevalence standard for rare diseases.

▪ Orphanet

Tot. Concepts = 7,715

Tot PT = 7,715

Tot. Syn = 5,224

Percentages of mapping MIM/ICD10
(concept) level):

- OMIM = 44.16%

- ICD10 = 32,30%

The so-called 'rare diseases' are diseases that affect a small number of people compared to the general population. **In Europe, a disease is considered as rare when it affects 1 person per 2,000.** However, this status may vary with time and also depends upon the area considered.

▪ NORD

Tot. Concepts = 1,236

Tot PT = 1,236

Tot Syn. = 4,562

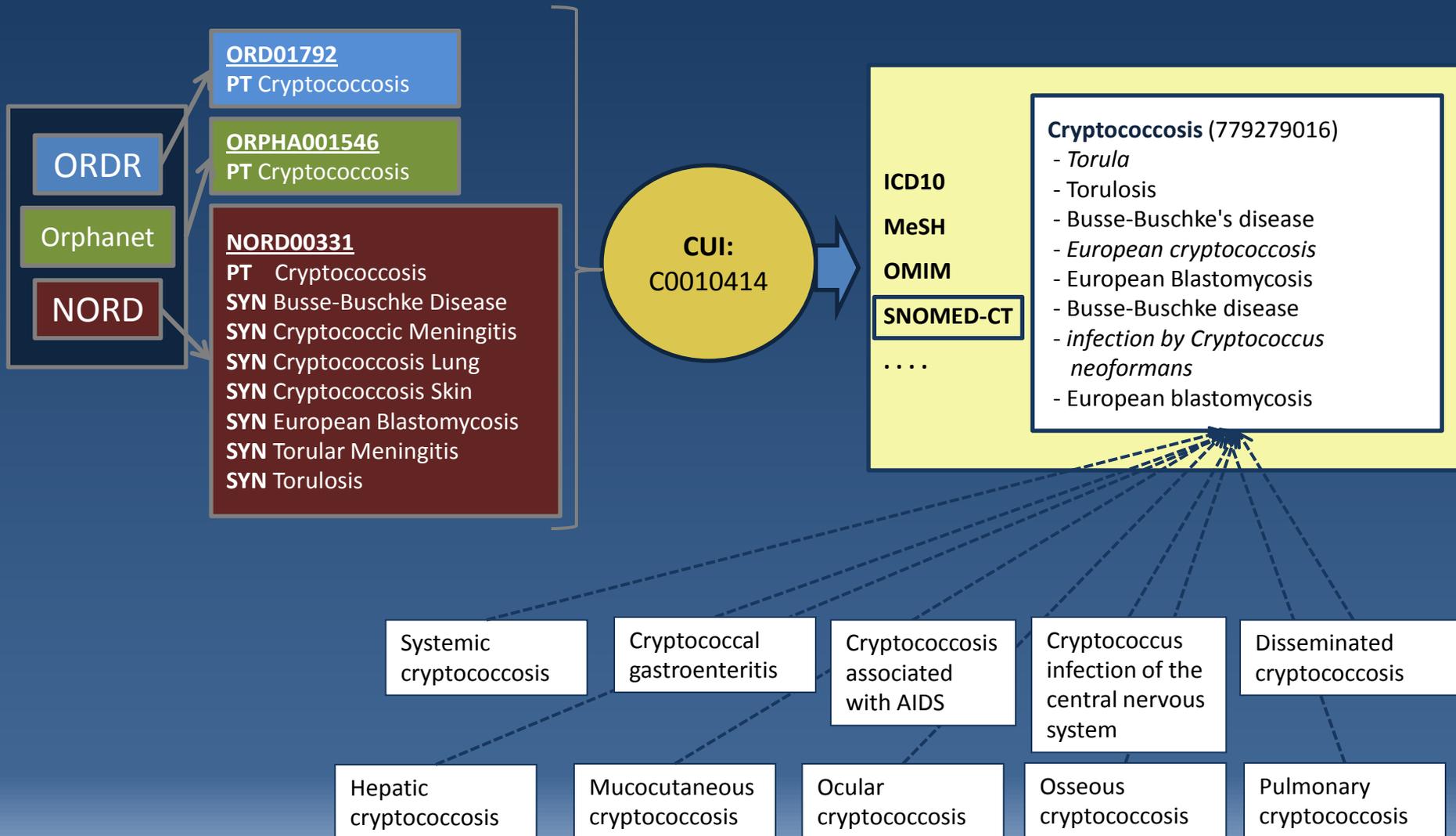
Disorder Subdivision = 1,283

NORD uses the same definition that the U.S. Food and Drug Administration uses: A rare disease is one that affects fewer than 200,000 Americans at any given time.

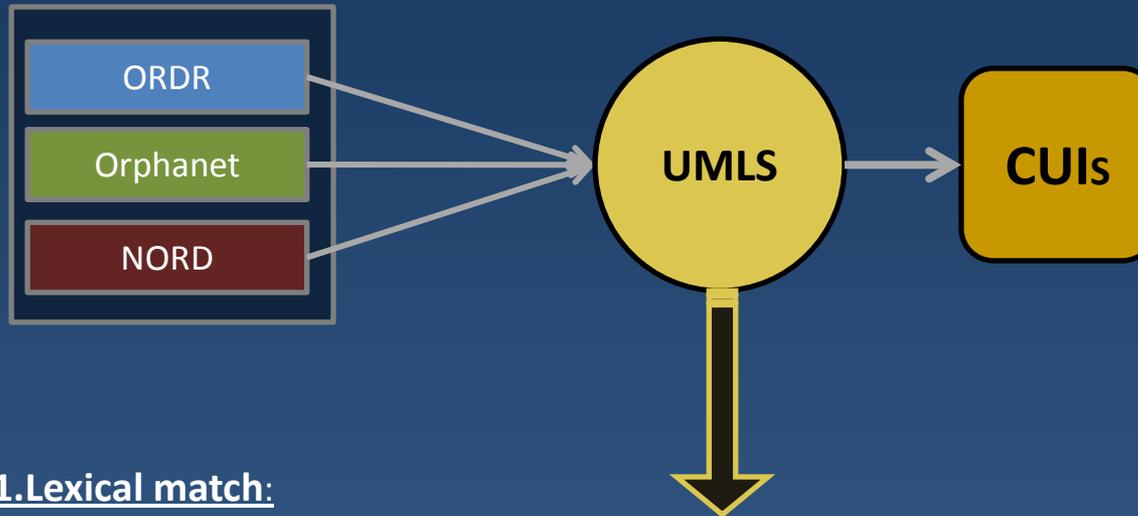
▪ UMLS (Unified Medical Language System)

2010AB

Implementation: overview



Methods: The mapping



1. Lexical match:

Glycogen storage disease type 4	C0017923	Exact Match
Isolated growth hormone deficiency type IA	C1849790	Normalized string

2. Validation:

- Semantic group filtering (Disorders):

Andersen disease	C0017923	Disorders
Andersen disease	C1415001	Genes & Molecular Sequences

- Manual review

Overlap among sources and representation in target sources

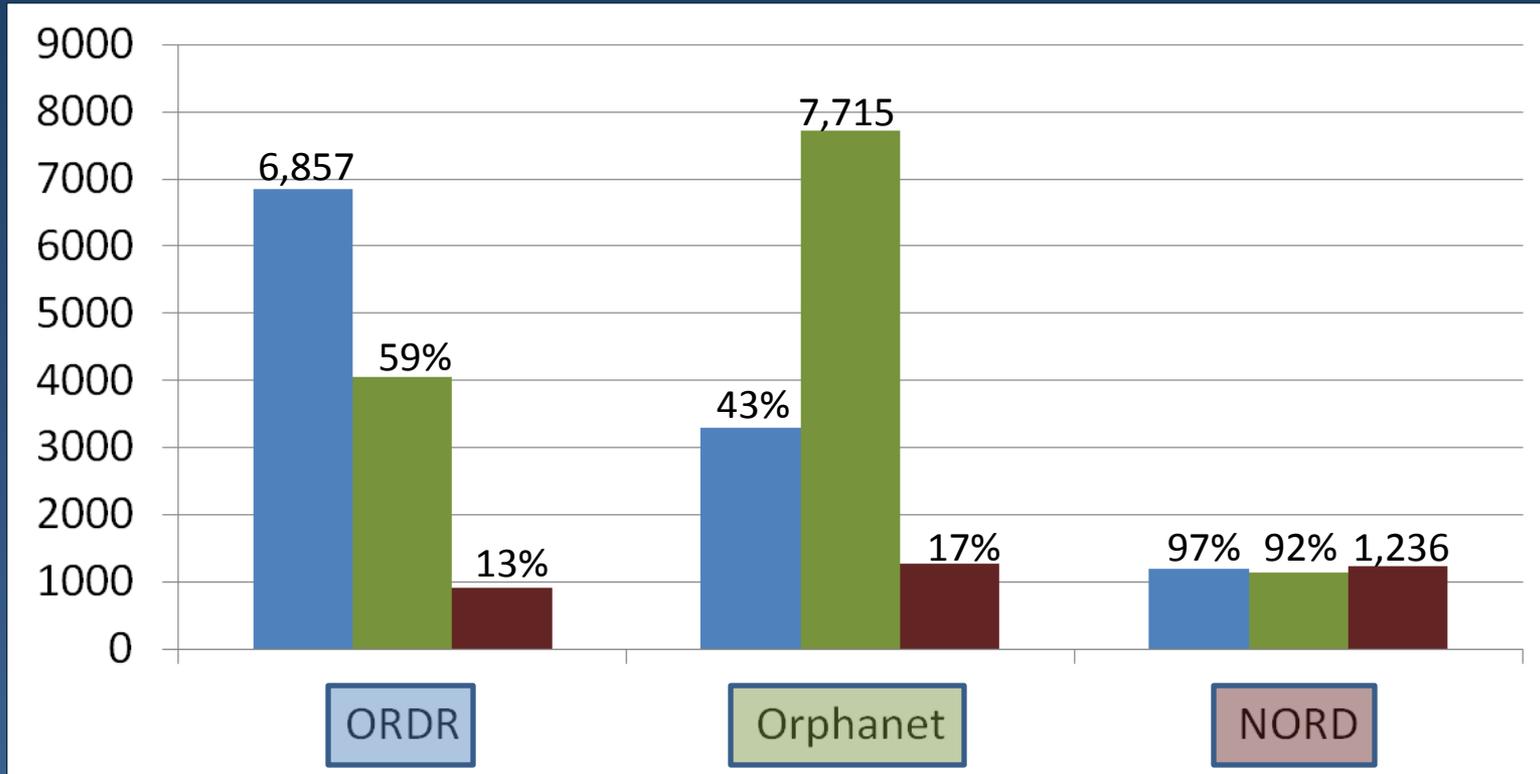
CUI	ORPHANET			CCS	CPM	ICD10	ICPC2P	JABL	KCD5	LNC	MDR	MSH	MSHLAV	MTH	MTHICD9	NAN	NCI	OMIM	OMs	SNOMEDCT	ULT	UWDA
	ORPHANET	NORD	ORD																			
	4400	2567	6250	69	1	1029	626	449	1028	101	2624	5663	3	5	1212	1	2039	3802	6	4192	1	3
C0000744	1	1	1									1			1		1	1		1		
C0000833		1	1				1			1	1	1			1					1		
C0000880	1										1	1					1			1		
C0000889	1	1	1			1			1		1	1			1		1	1		1		
C0001079	1	1	1			1			1		1	1			1		1			1		
C0001080	1	1	1			1	1		1		1	1			1		1	1		1		
C0001126			1								1	1					1	1		1		
C0001175		1					1				1	1			1		1			1		
C0001193	1	1	1								1	1			1			1		1		
C0001197			1								1	1					1			1		
C0001206	1	1	1				1				1	1					1	1		1		
C0001231		1	1			1			1		1	1			1		1			1		
C0001261			1			1			1		1	1			1		1			1		
C0001403	1	1	1			1	1		1		1	1			1		1	1		1		
C0001429			1								1	1					1			1		
C0001519		1	1								1	1			1		1	1		1		
C0001529	1	1	1								1	1					1	1		1		
C0001622	1		1								1	1			1			1		1		
C0001623			1				1				1	1					1	1		1		
C0001624	1										1	1					1			1		
C0001627	1		1						1		1	1			1		1	1		1		
C0001733			1								1	1			1		1	1		1		
C0001768	1	1									1	1			1			1		1		
C0001815	1	1	1								1	1			1		1	1		1		
C0001816			1			1			1		1	1			1		1	1		1		
C0001824			1			1	1		1		1	1			1		1	1		1		

Presence/absence of the concepts in the sources

Presence/ Absence of the Concepts in the target sources in the UMLS

Total number of UMLS concepts mapped to: 8,435

Overlap among sources



Representation in target sources

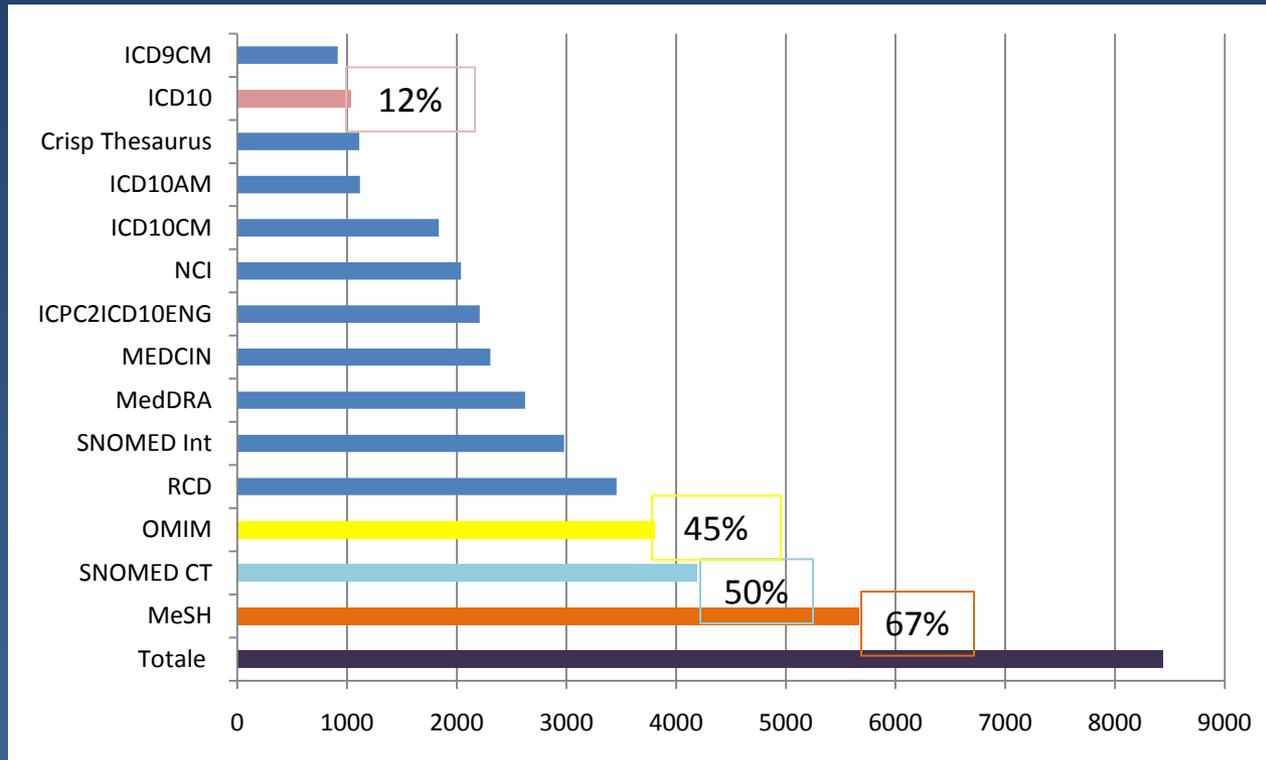
➤ Sources that have no representation

▪ Specific context:

- Foundational Model of Anatomy Ontology (FMA)
- NANDA nursing diagnoses: definitions & classification, 2004 (NAN)

Representation in target sources (2)

- Sources that have a good representation
(total number of UMLS concepts mapped to: 8,435)



Mapping to UMLS

	Unambiguous	Ambiguous	Unmapped	Total Concepts
ORDR	71%	7%	22%	6,857
Orphanet	49%	4%	47%	7,715
NORD	74%	24%	2%	1,236

Tab.1 Percentages of the Mapping to UMLS

1. **Unambiguous Concept**: The majority of terms of a given concept map to the same CUI
2. **Ambiguous Concept**: The terms of a given concept map several CUIs
3. **Unmapped**: No term of a given concept maps any Disorders in UMLS

Unambiguous mapping

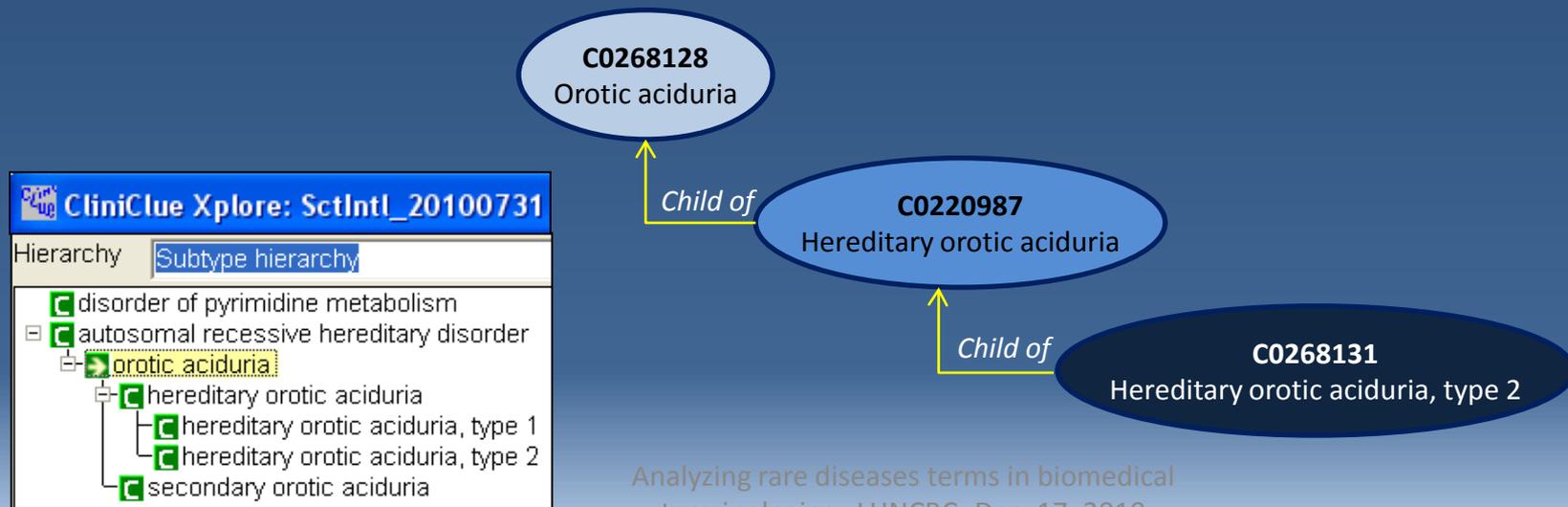
Code Source	Source Preferred Name	CUI	Source Preferred Name
ORD00117	Acrodysostosis	C0220659	Acrodysostosis
ORPHA001248	Maxillo-nasal dysplasia	C0220692	MAXILLONASAL DYSPLASIA, BINDER TYPE
NORD00312	Conn Syndrome	C1384514	Conn Syndrome

Tab.2 Examples of unambiguous mapping, the majority of terms of a given concept map to the same CUI

Ambiguous mapping related to granularity issues

ORPHA000030	CUI 1 C0268128	CUI 2 C0220987	CUI 3 C0268131
Oroticaciduria	Orotic aciduria		
Orotic aciduria hereditary		Hereditary orotic aciduria	
Orotidylic decarboxylase deficiency			Hereditary orotic aciduria, type 2
Uridine monophosphate synthetase deficiency	---	---	---

Tab3. Examples of Ambiguous Mapping: the terms of a given concept map to several CUIs



Ambiguous mapping not related to granularity

ORPHA000016	CUI1 C0339537	CUI2 C1844778
Blue cone monochromatism	Blue cone monochromatism	
Achromatopsia incomplete, X-linked		Achromatopsia, incomplete, x-linked
Achromatopsia, atypical, X linked	---	---
S-cone monochromatism	---	---

Tab4. Examples of ambiguous concept: the terms of a given concept map to several CUIs

Examples of Unmapped Concepts

Source Code	Full Name	Possible Explanation
ORPHA002096 ORDR05685	Richieri Costa-Guion Almeida-Cohen syndrome	Extremely rare disease
ORPHA002369	Lateral body wall complex	Extremely rare disease
ORD03770 NORD00694	Levy-Yeboah Syndrome	Recently recognized disorder (June 2006)

Tab5. Examples of unmapped concepts: no term of a given concept maps any Disorders in UMLS

Limitations

- Terms/concepts not found in UMLS:
 - Extremely rare diseases (Ex. *Lateral body wall complex*)
 - Recently recognized disorders (Ex. *Levy-Yeboa Syndrome*)
- Ambiguity related to granularity:
 - Different sources group terms in different ways
- Missing hierarchical information for Orphanet:
 - Overestimation of unmatched concepts (grouper vs. leaf concepts)
 - Grouper concepts : rare genetic skin disease
 - Leaf concepts: xeroderma pigmentosum
 - No ontological consistency validation possible with other sources

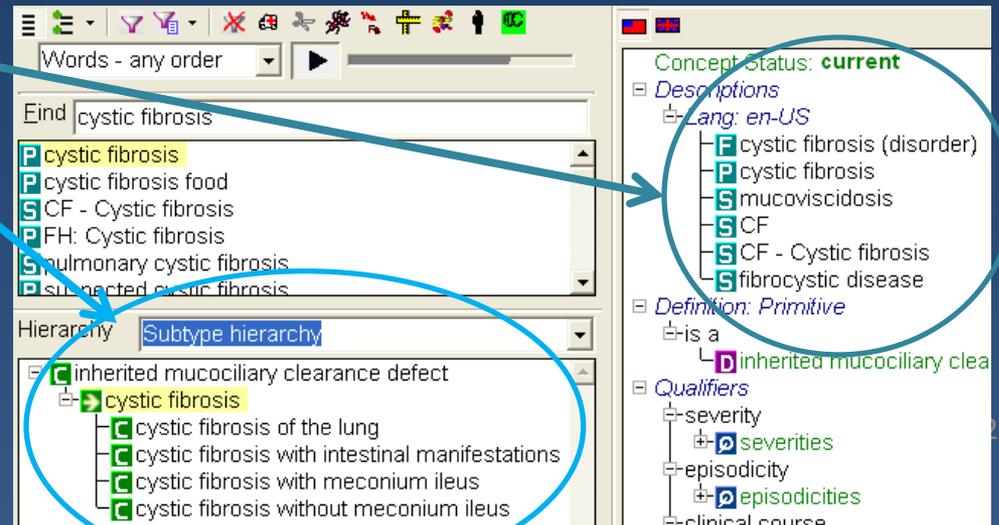
Unmapped	
ORDR	22%
Orphanet	47%
NORD	2%

Future work

➤ Looking for more specific information:

- Enhance information retrieval

- Additional synonyms
- Descendant concepts



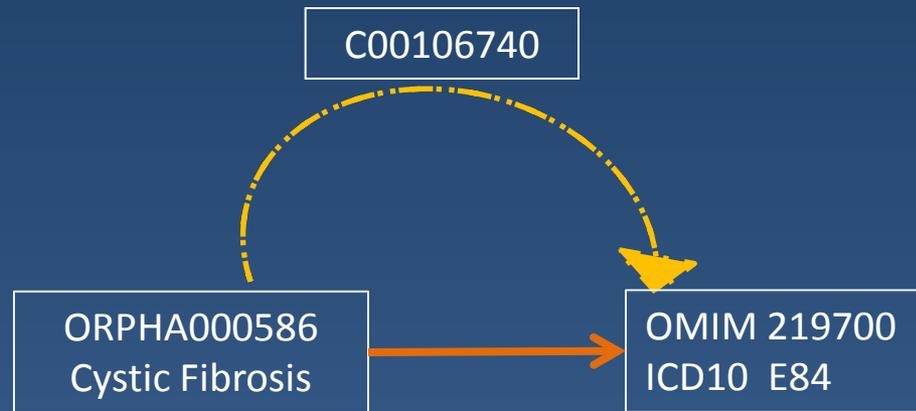
➤ Feedback to UMLS:

- Possible missed synonymy or missing hierarchical relation

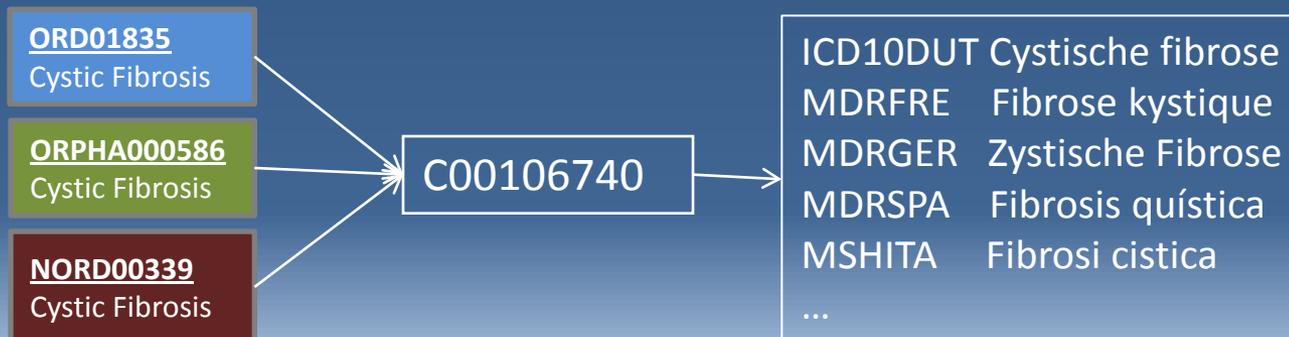
Full name ORPHA000016	CUI1 C0339537	CUI2 C1844778
Blue cone monochromatism	Blue cone monochromatism	
Achromatopsia incomplete, X-linked		Achromatopsia, incomplete, x-linked

Future work (2)

- For quality assurance compare:
 - Orphanet direct mapping to OMIM and ICD-10
 - Indirect correspondences found through the UMLS



- Coverage of rare disease terms in non- English languages



Acknowledgements

Dr. Olivier Bodenreider

Dr. Bastien Rance

Dr. Stuart J. Nelson

ORDR (Office of Rare Diseases Research)

Orphanet

NORD (National Organization for Rare Disorders)

Dr. Clement J. McDonald

Dr. Aaron B. Navarro

May Cheh

ORAU (Oak Ridge Associated Universities)

Thanks !



erika.pasceri@nih.gov