

On the Integration of Biomedical Databases: a Naive Approach

José Luis Oliveira
Univ. Aveiro, IEETA, PT
jlo@ua.pt

Barcelona, 27-06-2007



Objectives

- Integration of clinical and genetic information from public databases
- Focus on rare diseases
- Why rare diseases?
 - ~80% do have genetic association
 - ~ 6,000 distinct rare diseases
 - 6% to 8% of the population

Main Questions: what are the ... ?

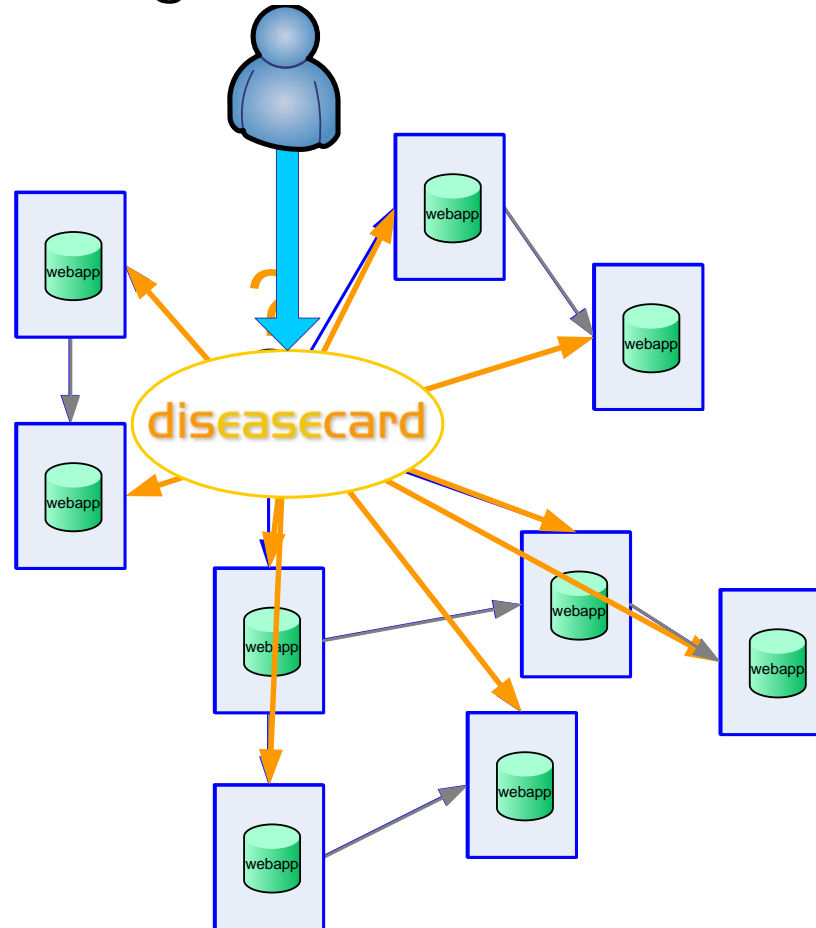
- **symptoms** of the disease?
- main **publications** about this disease?
- **drugs** available for the disease?
- gene **therapies** or **clinical trials** for the disease?
- **laboratories** that perform genetic tests?
- **genes** that cause the disease?
- **chromosomes** for these genes?
- **mutations** found in these genes?
- **names** used to refer to these genes?
- **proteins** coded by these genes?
- the **functions** of the gene product?
- **3D structure** for these proteins?
- **enzymes** associated to these proteins?

Public Resources

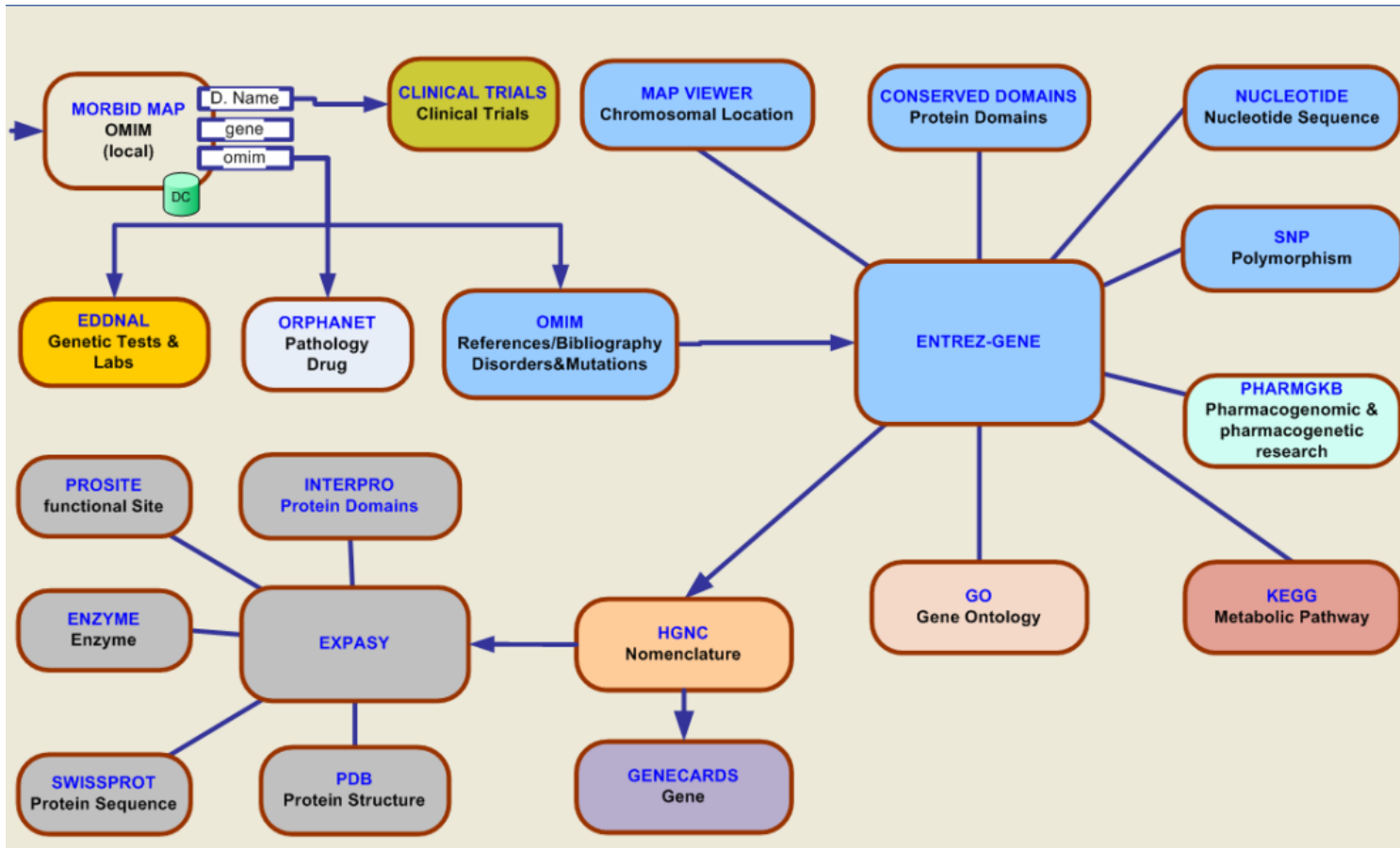
- Orphanet
- IIER
- ClinicalTrials
- OMIM
- NCBI dbSNP
- EMBL Nucleotide Sequence Database
- Entrez-Gene
- Swiss-Prot
- ...
- ProDom
- KEGG
- GeneCards
- PubMed
- PharmGKB
- HGNC (HUGO)
- GO
- EDDNAL
- ...

Motivation

- Providing an intuitive interface to the main biomedical and genetic resources



Navigation Model



Methods

- Disease **cards**
 - Each card contains a set of biomedical concepts
- Navigation **protocol**
 - Described in a formal language (XML)
- A web **crawler** for each concept
 - The crawler specify where to search and what to retrieve
 - Used just for the construction of the card
 - Multithreading
- No physical integration!

diseasecard

Welcome to DiseaseCard Tool

>Search for a disease

by disease name

>[List of available disease](#)

- by disease name
- by omim number
- by gene symbol


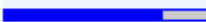











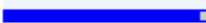


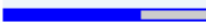




DiseaseCard is an information retrieval tool for accessing and integrating genetic and medical information for health applications. Resorting to this integrated environment, clinicians are able to access and relate diseases data already available in the Internet, scattered along multiple databases. Diseasecard was developed by [Bioinformatics Group of University of Aveiro](#).

The use of DiseaseCard is subject to the following [disclaimer and warning](#).

Endorsement:



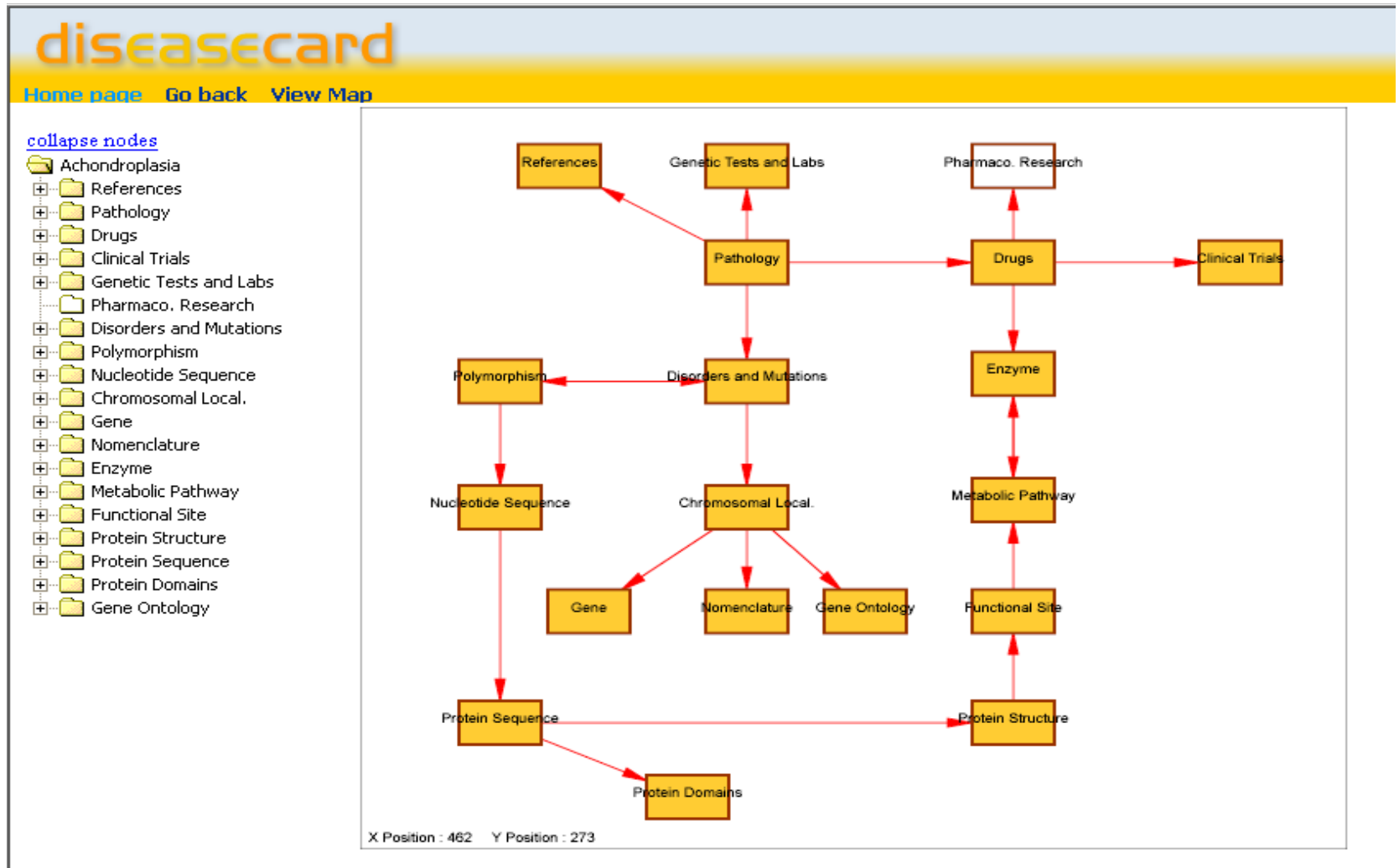
Diseases List

Diseasename	Synonyms	Coverage	Last Update
Aarskog-Scott syndrome		 14/19	2007-01-23
ABCD syndrome			06-09-27
Abdominal obesity-metabolic syndrome			06-10-30
Abetalipoproteinemia			07-02-04
Acampomelic campolelic dysplasia			06-09-29
ACAT2 deficiency			07-02-07
Acatlasemia			07-02-07
Acetyl-CoA carboxylase deficiency			06-09-27
Achalasia-addisonianism-alacrimia syndrome			07-02-04
Acheiropody		 11/19	2006-10-25
Achondrogenesis Ib		 14/19	2007-02-04
Achondrogenesis-hypochondrogenesis, type II		 15/19	2006-06-06
Achondroplasia		 17/19	2006-10-24
Achromatopsia-1		 5/19	2006-05-23
Achromatopsia-2		 12/19	2006-10-03
Achromatopsia-3		 12/19	2006-05-03
Achromatopsia-4		 9/19	2006-07-19
Acid-labile subunit, deficiency of		 10/19	2006-05-06

Synonyms

- Aarskog-scott syndrome ;
- Aas faciogenital dysplasia with attention deficit-hyperactivity disorder, included ;
- Faciodigitogenital syndrome ;
- Faciogenital dysplasia ;
- Fgdy ;

Navigation Protocol



Card Information

diseasecard
[Home page](#) [Go back](#) [View Map](#)

collapse nodes

- Achondroplasia
 - References
 - 134934
 - 100800
 - Pathology
 - Achondroplasia**
 - Drugs
 - Clinical Trials
 - Genetic Tests and Labs
 - Pharmaco. Research
 - Disorders and Mutations
 - 100800
 - Polymorphism
 - Nucleotide Sequence
 - Chromosomal Local.
 - Gene
 - Nomenclature
 - Enzyme
 - Metabolic Pathway
 - Functional Site
 - Protein Structure
 - Protein Sequence
 - Protein Domains
 - Gene Ontology

orphanet
 serveur d'information pour tous publics
 sur les maladies rares et les médicaments orphelins

← HOME SERVICES FOR PROFESSIONALS

Print

DISEASE : **Achondroplasia** ICDQ77.4

MIM: [100800](#)

[Scientific publications PubMed](#)

[Clinical signs\(15\)](#)

[Other website\(s\)\(12\)](#)

► **Outpatient clinic(s)**

- [Bone diseases clinic](#)
- [Dysmorphology clinic](#)
- [Genetic counselling clinic](#)

► [Diagnostic laboratory \(58\)](#)

► [Research projects \(8\)](#)

► [Networks \(6\)](#)

► [Support group\(s\) \(13\)](#)

► [Registries / Observatories \(2\)](#)

► [Drugs \(1\)](#)

About Orphanet ■
 About rare diseases ■
 About orphan drugs ■
 Support groups ■

Warning !

Headquarter •
 Contact •
 Order the Orphanet book •

Home | Help | Contact | Quality Charter | Statistics

Local intranet

Disease Forum

diseasecard

home back root help

FORUM

[collapse nodes](#)

- [-] {Alzheimer disease 9, late onset, s
- [-] References
- [-] Pathology
- [-] Drugs
- [-] Clinical Trials
- [-] Genetic Tests and Labs
- [-] Pharmaco. Research
- [-] Disorders and Mutations
- [-] Polymorphism
- [-] Nucleotide Sequence
- [-] Chromosomal Local.
- [-] Gene
- [-] Nomenclature
- [-] Enzyme
- [-] Metabolic Pathway
- [-] Functional Site
- [-] Protein Structure
- [-] Protein Sequence
- [-] Protein Domains
- [-] Gene Ontology

Forum for **{Alzheimer disease 9, late onset, susceptibility to}**

This is a space where users can share their experience on [this particular disease](#). Do you know other important sources of information? Do you want to share relevant ad You are welcome to share your experience.

page [\[1\]](#) [Post Mess](#)

Title: Usefull link Submitted by: hImp 2007-02-08 15:05


This Link is usefull to understand what happens in the alzheimer disease

<http://www.diseases-explained.com/Alzheimers/whatcausesalzhei.html>

Signature:Good Day

Title: This image illustrates the alzheimer disease Submitted by: hImp 2007-02-08 15:02

Brain Cross-Sections



Sulcus Sulcus

Gyrus Gyrus

Conclusions

- Portal for rare diseases www.diseasecard.org
- Navigation from the symptom to gene
 - Integrated view
- Accurate information
 - Well known databases, continuously updates
- No need to master biological databases
 - Avoids dealing with search boxes, nomenclatures, etc.
- Information retrieval engine based on web wrappers
 - Easy configuration

- The model can be applied to other application fields!

Credits

- Univ. Aveiro, PT
 - Gaspar Dias
 - Hugo Pais
 - José Luis Oliveira
- ISCIIII, SP
 - Fernando Martin-Sanchez
 - Javier Vicente

www.diseasecard.org

<http://bioinformatics.ua.pt>