Knowledge Bases and Terminology
S81 – November 11, 2008

A Framework for Characterizing Drug Information Sources

Mark Sharp, Nina Wacholder
SCILS, Rutgers University
New Brunswick, New Jersey

Olivier Bodenreider
National Library of Medicine
Bethesda, Maryland

Motivation

- Many drug information sources available

MedlinePlus
Trusted Health Information for You

ChEBI

Drugs@FDA
FDA Approved Drug Products

KEGG DRUG

DailyMed
Current Medication Information

MeSH

ChemIDplus Advanced

PubChem Substance

DrugBank

Motivation

- Many use cases for drug information sources
  - E-prescribing
  - Clinical decision support
  - Patient information
  - Clinical/pre-clinical research
- Challenges
  - Few integration systems (RxNorm)
  - No registries of drug information sources
- Difficult to pair drug information sources with use cases

Objectives

- Framework for characterizing drug information sources
- Application to
  - Comparing sources
  - Selecting sources most relevant to a given use case

Materials

- 23 drug information sources selected
  - Publicly available
  - Electronic format
  - Explicit data elements
  - Coverage (quantitative)
  - Type of information covered
    - Pharmacy (e.g., trade name, manufacturer)
    - Chemistry (e.g., empirical formula, solubility)
    - Biology (e.g., molecular target, pathway)
    - Clinical medicine (e.g., indications, interactions)
  - Purposely excluded information
    - Pricing, packaging, retail and patent

Source Name	Website	# INs
MedMaster	http://www.nlm.nih.gov/medlineplus/druginformation.html	nd
DrugDigest	http://www.drugdigest.org/DD/Home	~1,000
DailyMed	http://dailymed.nlm.nih.gov	1,117
ClinicalTrials.gov	http://clinicaltrials.gov/	924
DrugInfo	http://druginfo.nlm.nih.gov/	">12,000"
RXNORM	http://mor.nlm.nih.gov/download/rxnav/	5,592
Drugs@FDA	http://www.accessdata.fda.gov/scripts/cder/drugsatfda/	1,689
WHO-ATC	http://www.whocc.no/atcddd/	~3,000
WHO-DRUG	http://www.umc-products.com/DynPage.aspx?id=2829&mn=1107	9,899
Int'l Pharm.	http://www.who.int/medicines/publications/pharmacopoeia/en/index.html	420
INN	http://www.who.int/medicines/services/inn/en/index.html	~2,000
USP Dictionary	http://www.uspusan.com/usan/login	>4,317
USAN via AMA	http://www.ama-assn.org/ama/pub/category/2956.html	689
MeSH MH only	http://www.nlm.nih.gov/mesh/meshhome.html	~2,000
MeSH all	http://www.nlm.nih.gov/mesh/meshhome.html	~5,000
UMLS	http://www.nlm.nih.gov/research/umls/	~9,000
PubChem	http://pubchem.ncbi.nlm.nih.gov/	nd
ChemiIDplus	http://chem.sis.nlm.nih.gov/chemidplus/	nd
ChEBI	http://www.ebi.ac.uk/chebi/	>7,000
DrugBank	http://redpoll.pharmacy.ualberta.ca/drugbank/	1,835
KEGG DRUG	http://www.genome.jp/kegg/drug/	6,848
Reactome	http://www.reactome.org/	nd
HumanCyc	http://humancyc.org/	20

Methods Building the framework (1/2)

- Establish a list of features in the 23 sources
  - Manually
  - Database schema / web page organization
  - Same query performed on multiple sources
- Normalize the features
  - "Dimensions of drug information"
  - E.g., for trade names
    - Data elements: Brand name, Trademark, …
    - Values: Propecia, Bayer Aspirin, Tylenol, …
- Group the dimensions by domain
  - Hierarchical facets

Matrix of Sources x Features

Domain	Dimension	Source: MedNames	DrugGuru	Dailymed	ClinicalTrials.gov	Drugbank	SDWCDM	Drugs@FDA	WHO ATC	WHO DDD	Int'l Pharm	RxV	GSK Consumer	MEDVAL vs AMA	MedPortal	MedWeb	MyLS	RxCore	ClinicalTrials	GRSS	DrugBank	RXNORM/UMLS	Stanford	Hanover
Business	trade names	■	■	■	○	■	■	■	■	■				■	■	■	■	■			■			
	distributors		○	■	○		■	■	○	○						■	■			○	■	○		
	combo products	■	■	■	■	■	■	■	■				■					■	■	■		■	■	
	manufacturers			■			■						■	■										
	control drug class												■	■	■	■	■	■	■					
	approval info							■													■			
Chemistry	chemical class			■							■	■	■	■	■	■	■	■	■	■	■			■
	CAS#										■		■	■	■	■	■	■	■	■	■	■		
	structure graphic			■							■	■	■	■				■	■	■	■	■		■
	empirical formula			■							■	■	■	■				■	■	■	■	■		■
	InChI																	■	■	■	■			
	SMILES																	■	■	■	■			■
	molecular structure																	■		■		■	■	
	H bond donors[a]																	■						
	H bond acceptors[a]										■							■						
	molecular weight[a]			■							■			■				■		■	■	■		
	solubility[a]			■							■							■	■	■				
	chem synonyms					■			■						■	■	■				■			
	physical class											■									■			
	mechanism											■							■		■			
	pKa											■									■			
	extra chemistry											■												
Biology	molecular targets			■						○								■			■	■	○	
	mech of action			■						○								■			■			
	biological effects			■						○								■			■			■
	metabolism			■														■			■		■	
	extra ADME			■																	■			
	anatomy																			■	■			
	taxonomy								■	■								■						
	pathways																	■						
	therapeutic class		■	■	■	■			■	■	■		■	■	■	■	■	■	■	■	■	■		■
Clinical	indications	■	■	■	■													■			■			
	contraindications	■	■	■														■			■			
	side effects/warnings	■	■	■														○						
	drug interactions	■	■	■														■			■			
	general info	■	■	■																	■			
	research			■																	■			
	in progress/approvals			■										○				■			■			

Methods Building the framework (2/2)

- Evaluation of quantitative coverage
  - Number of single-component approved generic names
- Other elements assessed
  - Source availability
    - Cost, license
  - Update frequency
  - Existence of an application programming interface
  - Format

Methods Evaluating the framework

- Grouping sources
  - Correspondence analysis on the matrix of Sources x Features (MSVP)
  - Sources with the same profile of features are grouped together
- Selecting sources
  - Hypothetical use case
  - Use the matrix to select sources for this use case

Domain	Characteristic	Source:	MedNames	DrugBank	DailyMed	ClinicalTrials.gov	DrugInfo	RxNorm/RxN	DrugBPDA	WHO-ATC	WHO-DDD	Int'l Pharm	RxV	EBP/Consumer	KEGG/Ver.ANA	MedRxiv	MedPrint	UptLG	PubChem	ChemTaxo	CHEBI	DrugBank	ISSN DRUG	Rxnorm	HumanCyc
Taxonomy	usel names	■	■	■	○	■	■	■	■	■	■			■	■	■	■	■	■	■	■	■	○	○	
	dosa forms		○	■	○		■	■	○	○							■	■				○	○	■	
	combo products	■	■	■	■	■	■	■	■					■			■	■	■	■		■	■		
	manufacturers			■				■						■	■										
	name/ code assoc													■	■	■	■	■	■	■					
	approval info									■												■			
Ontology	chemical assoc			■								■	■	■	■	■	■	■	■	■	■	■			■
	CAS#											■		■	■	■	■	■	■	■	■	■			
	structure graphic			■								■	■	■	■				■	■	■	■	■		■
	computational formula			■								■	■	■	■				■	■	■	■	■		■
	InChI[1]																		■	■	■	■			
	SMILES[1]																		■	■	■	■			■
	molec structure																		■	■			■	■	
	H bond donors[3]																		■						
	H bond acceptors[3]																		■						
	molecular weight[3]			■								■			■				■		■	■	■		
	solubility[3]			■								■							■	■	■	■			
	chem synonyms					■			■							■	■	■				■			
	physical chem												■										■		
	exchange prop												■								■		■		
	pKa												■										■		
	extra chemistry												■						■	■	■	■	■	■	■
Biology	molecular targets					■					○								■				■	○	○
	mech of action					■					○								■				■		
	biological effects					■					○								■				■		■
	metabolism					■													■				■	■	
	extra ADME					■																	■		
	toxicity																				■		■		
	anatomy									■	■								■						
	histology																				■				
	pathways																								■
Clinical	therapeutic effects		■	■	■	■			■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	
	indications	■	■	■	■														■				■		
	contraindications	■	■	■															■				■		
	side effects/warns	■	■	■															○						
	drug interactions	■	■	■															■						
	patient info	■	■	■																			■		
	research			■																			■		
	organ/cell tgt[3]				■										○				■				■		

Results Framework

Pharmacy
- trade names
- dose/form
- combo products
- manufacturer
- manufacturer code name
- approval info.

Chemistry
- chemical name
- CAS#
- structure graphic
- empirical formula
- InChI
- SMILES
- similar structures
- H bond donors
- H bond acceptors
- molecular weight
- solubility
- chem. superclass
- physical descr.
- melting point
- pKa
- other chemistry

Biology
- molecular target
- mech. of action
- biological effect
- metabolism
- other ADME
- toxicity
- anatomy
- bioassay
- pathways

Clinical
- therapeutic class
- indication
- contraindication
- side eff/prec/warn
- drug interactions
- patient info
- research lit.
- experimental app's

- 39 dimensions organized in 4 groups

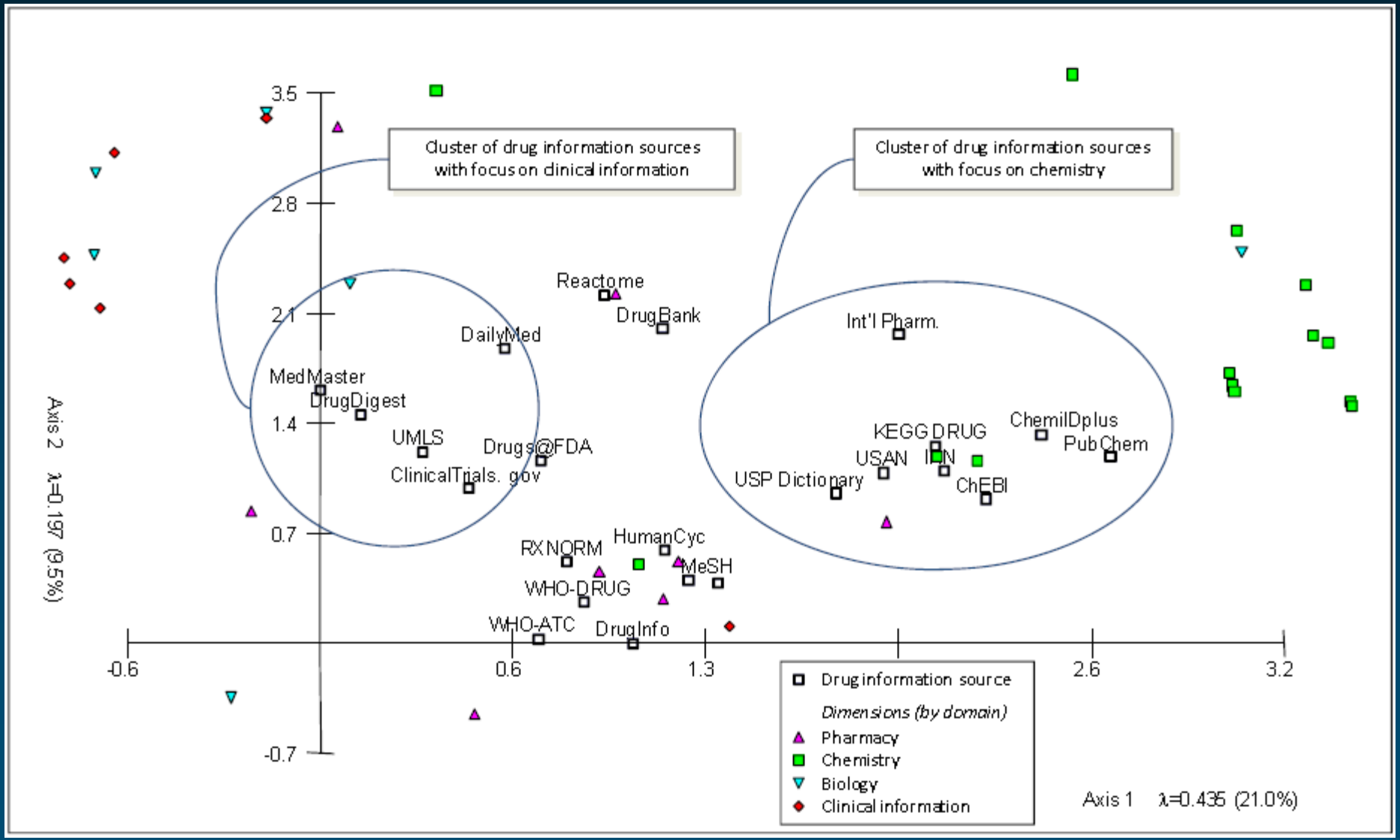
Results Matrix of Sources x Features

Domain	Dimension / Source	MedMaster	DrugDigest	DailyMed	ClinicalTrials.gov	DrugInfo	RXNORM	Drugs@FDA	WHO-ATC	WHO-DRUG	Int'l Pharm.	INN	USP Dictionary	USAN via AMA	MeSH MH	MeSH all	UMLS	PubChem	ChemIDplus	ChEBI	DrugBank	KEGG DRUG	Reactome	HumanCyc
pharmacy	trade names	●	●	●	○	●	●	●		●			●	●	●	●	●	●	●	●	●	○	○	
	dose/form		○	●	○		●	●	○	○						●	●				○			
	combo products	●	●	●	●	●	●	●	●	●			●				●	●	●		●	●		
	manufacturer			●				●					●	●										
	manuf. code name												●	●	●	●	●	●	●					
	approval info.							●													●			
chemistry	chemical name			●							●	●	●	●	●	●	●	●	●	●	●			●
	CAS#										●		●	●	●	●	●	●	●	●	●	●		
	structure graphic			●							●	●	●	●				●	●	●	●	●		●
	empirical formula			●							●	●	●	●				●	●	●	●	●		●
	InChI[4]																	●	●	●	●			
	SMILES[5]																	●	●	●	●			●
	similar structures																	●	●		●	●		
	H bond donors[6]																	●						
	H bond acceptors[6]																	●						
	molecular weight[6]			●							●			●				●		●	●	●		
	solubility[6]			●							●							●	●		●			
	chem. superclass					●			●						●	●	●			●				
	physical descr.										●										●			
	melting point										●								●		●			
	pKa										●										●			
	other chemistry										●							●	●	●	●	●	●	

Domain	Dimension / Source	MedMaster	DrugDigest	DailyMed	ClinicalTrials. gov	DrugInfo	RXNORM	Drugs@FDA	WHO-ATC	WHO-DRUG	Int'l Pharm.	INN	USP Dictionary	USAN via AMA	MeSH MH	MeSH all	UMLS	PubChem	ChemIDplus	ChEBI	DrugBank	KEGG DRUG	Reactome	HumanCyc
pharmacy	trade names	●	●	●	○	●	●	●		●			●	●	●	●	●	●	●	●	●	○	○	
	dose/form		○	●	○		●	●	○	○						●	●				○			
	combo products	●	●	●	●	●	●	●	●	●			●				●	●	●		●	●		
	manufacturer			●				●					●	●										
	manuf. code name												●	●	●	●	●	●	●					
	approval info.							●													●			
chemistry	chemical name			●							●	●	●	●	●	●	●	●	●	●	●			●
	CAS#										●		●	●	●	●	●	●	●	●	●	●		
	structure graphic			●							●	●	●	●				●	●	●	●	●		●
	empirical formula			●							●	●	●	●				●	●	●	●	●		●
	InChI[4]																	●	●	●	●			
	SMILES[5]																	●	●	●	●			●
	similar structures																	●	●		●	●		
	H bond donors[6]																	●						
	H bond acceptors[6]																	●						
	molecular weight[6]			●							●			●				●		●	●	●		
	solubility[6]			●							●							●	●		●			
	chem. superclass					●			●						●	●	●			●				
	physical descr.										●										●			
	melting point										●								●		●			
	pKa										●										●			
	other chemistry										●							●	●	●	●	●	●	

Domain	Dimension / Source	MedMaster	DrugDigest	DailyMed	ClinicalTrials. gov	DrugInfo	RXNORM	Drugs@FDA	WHO-ATC	WHO-DRUG	Int'l Pharm.	INN	USP Dictionary	USAN via AMA	MeSH MH	MeSH all	UMLS	PubChem	ChemIDplus	ChEBI	DrugBank	KEGG DRUG	Reactome	HumanCyc
biology	molecular target			●					○								●				●	●	○	
	mech. of action			●					○								●				●			
	biological effect			●					○								●				●			●
	metabolism			●													●				●		●	
	other ADME			●																	●			
	toxicity																		●		●			
	anatomy								●	●							●							
	bioassay																	●						
	pathways																						●	
clinical	therapeutic class		●	●	●	●			●	●	●		●	●	●	●	●	●	●	●	●	●		
	indication	●	●	●	●												●				●			
	contraindication	●	●	●													●				●			
	side eff/prec/warn	●	●	●													○							
	drug interactions	●	●	●													●				●			
	patient info	●	●	●																	●			
	research lit.			●																	●			
	experimental app's				●									○			●				●			

Results Grouping sources (PCA)



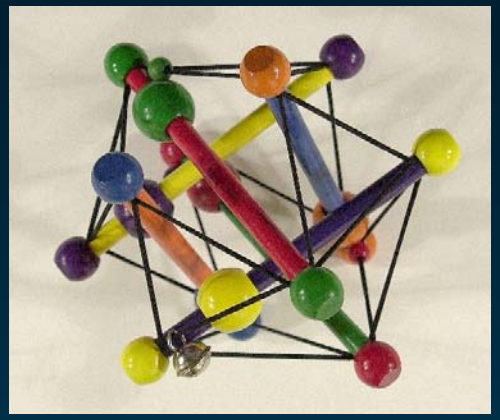
Results Selecting sources

- Use case
  - Find the drugs corresponding to a given indication or vice versa
  - Requirements
    - At least: *drug names*, *indications*
    - Additionally: *therapeutic class*, *mechanism of action*, *biological effect*, *molecular target*, *experimental applications*, and *chemical superclass*
- Best sources
  - UMLS, DrugBank, DailyMed, WHO ATC

Domain	Dimension / Source	MedMaster	DrugDigest	DailyMed	ClinicalTrials. gov	DrugInfo	RXNORM	Drugs@FDA	WHO-ATC	WHO-DRUG	Int'l Pharm.	INN	USP Dictionary	USAN via AMA	MeSH MH	MeSH all	UMLS	PubChem	ChemIDplus	ChEBI	DrugBank	KEGG DRUG	Reactome	HumanCyc
pharmacy	trade names	●	●	●	○	●	●	●		●			●	●	●	●	●	●	●	●	●	○	○	
	dose/form		○	●	○		●	●	○	○						●	●				○			
	combo products	●	●	●	●	●	●	●	●	●			●				●	●	●		●	●		
	manufacturer			●				●					●	●										
	manuf. code name												●	●	●	●	●	●	●					
	approval info.							●													●			
biology	molecular target			●					○								●				●	●	○	
	mech. of action			●					○								●				●			
	biological effect			●					○								●				●			●
	metabolism			●													●				●		●	
	other ADME			●																	●			
	toxicity																		●		●			
	anatomy								●	●							●							
	bioassay																	●						
	pathways																						●	
clinical	therapeutic class		●	●	●	●			●	●	●		●	●	●	●	●	●	●	●	●	●		
	indication	●	●	●	●												●				●			
	contraindication	●	●	●													●				●			
	side eff/prec/warn	●	●	●													○							
	drug interactions	●	●	●													●				●			
	patient info	●	●	●																	●			
	research lit.			●																	●			
	experimental app's				●									○			●				●			

Summary

- 23 drug information sources selected and analyzed
- 39 features (dimensions) extracted
  - Organized into 4 major categories
- Dimensions of drug information provide a framework for analyzing drug information sources
- Applications
  - Grouping sources
  - Matching sources with use cases



Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA