biowisdom
Part of the Instem Group

Repurposing of drug label information to create actionable intelligence with BioWisdom's Metawise

Paul Bradley

12th August, 2011

BioWisdom © 2011

Background

- Toxicity is now the major cause of drug failure
- Strong focus on developing predictive methods
  - Fail early, avoid progressing toxic candidates
- Improved prediction requires best view of current and *historic* data
  - Current emphasis on repurposing legacy data
  - Economic advantages, avoid costly repetition
- Harnessing legacy data can be a big problem
  - Different locations
  - Lack of structure, i.e. free text
  - Heterogeneity, lack of terminology standards
- Impedes alignment and analysis
- Difficult to answer simple questions
  - Which drugs elevate ALT level?

BioWisdom © 2011

Dealing with Unstructured Text

- Impracticable to manually review and extract data from unstructured sources
  - Automated methods needed to augment the process
  - Employ costly curation judiciously
- Conventional text analytics has limitations
  - NLP and entity recognition leave us with a classification problem
  - Tailoring to a specific domain, e.g. life sciences
  - Dictionary look-ups require large quantities of "synonyms"
    - Keeping current is a burden
- Need something more flexible
  - Tolerate linguistic variation
  - Easier to update
  - Facilitate adoption of standards
    - Harmonise legacy data with current research

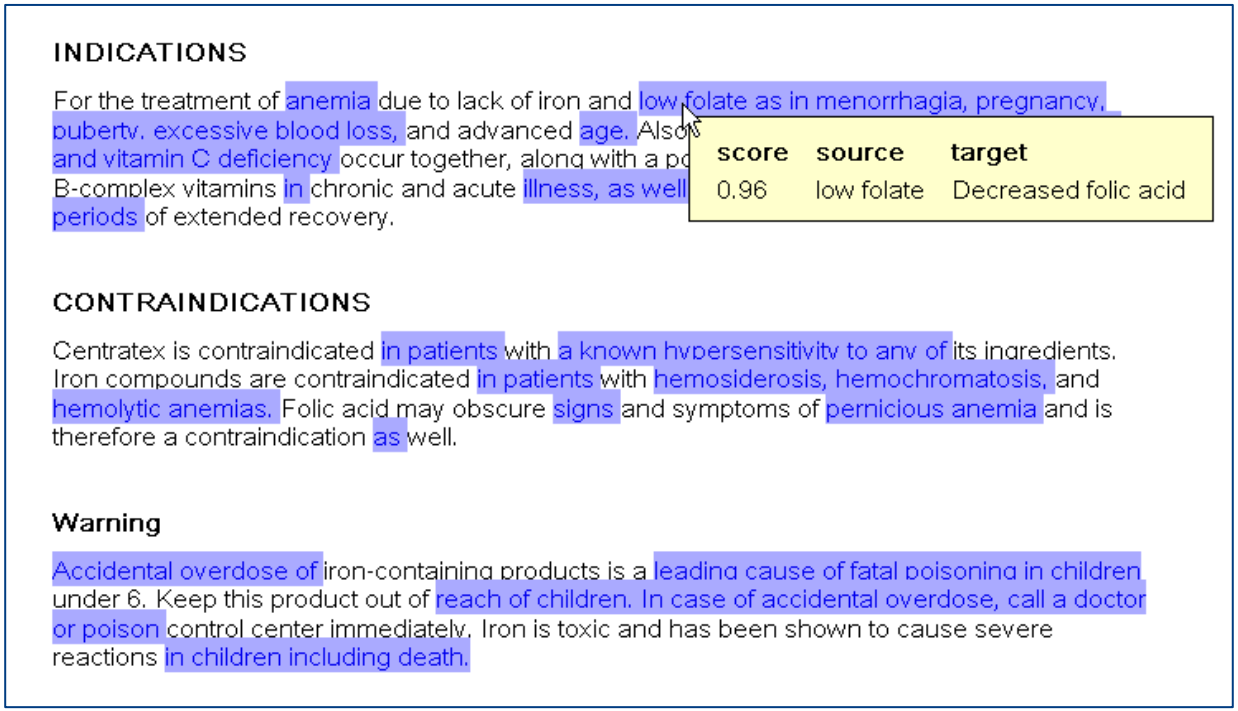
BioWisdom © 2011

BioWisdom's Metawise

- A terminology management system
  - Applications in enterprise search, literature searching, document mark up, term translation/harmonisation
  - More powerful than conventional, static dictionary-based methods
  - Robust, can deal with the linguistic diversity that exists in free text
- Template matching/extraction: recognition of important concepts based on term structure
  - Utilises term *categories*, clusters of replaceable terms, to create *templates*
  - hepatic necrosis - organ_location:process - pancreatic apoptosis
  - Identifies terms that are similar to known terms
- Term translation: convert free text references to any standard, e.g. SEND, SNOMED CT, ICD-10
  - Exhaustive matching: word substitutions, rearrangements, stemming
  - Translate/map new terms (e.g. from template matching) to a target ontology
  - Map between ontologies and standards

BioWisdom © 2011

Entity Recognition and Term Translation



INDICATIONS

For the treatment of anemia due to lack of iron and low folate as in menorrhagia, pregnancy, puberty, excessive blood loss, and advanced age. Also and vitamin C deficiency occur together, along with a po B-complex vitamins in chronic and acute illness, as well periods of extended recovery.

score	source	target
0.96	low folate	Decreased folic acid

CONTRAINDICATIONS

Centratex is contraindicated in patients with a known hypersensitivity to any of its ingredients. Iron compounds are contraindicated in patients with hemosiderosis, hemochromatosis, and hemolytic anemias. Folic acid may obscure signs and symptoms of pernicious anemia and is therefore a contraindication as well.

Warning

Accidental overdose of iron-containing products is a leading cause of fatal poisoning in children under 6. Keep this product out of reach of children. In case of accidental overdose, call a doctor or poison control center immediately. Iron is toxic and has been shown to cause severe reactions in children including death.

Candidate Term

low folate

Synonyms & Metadata

```
low    : syn : decreased
low    : syn : reduced
folate : syn : folic acid
folate : syn : vitamin b9
```

Target Translation

Decreased folic acid

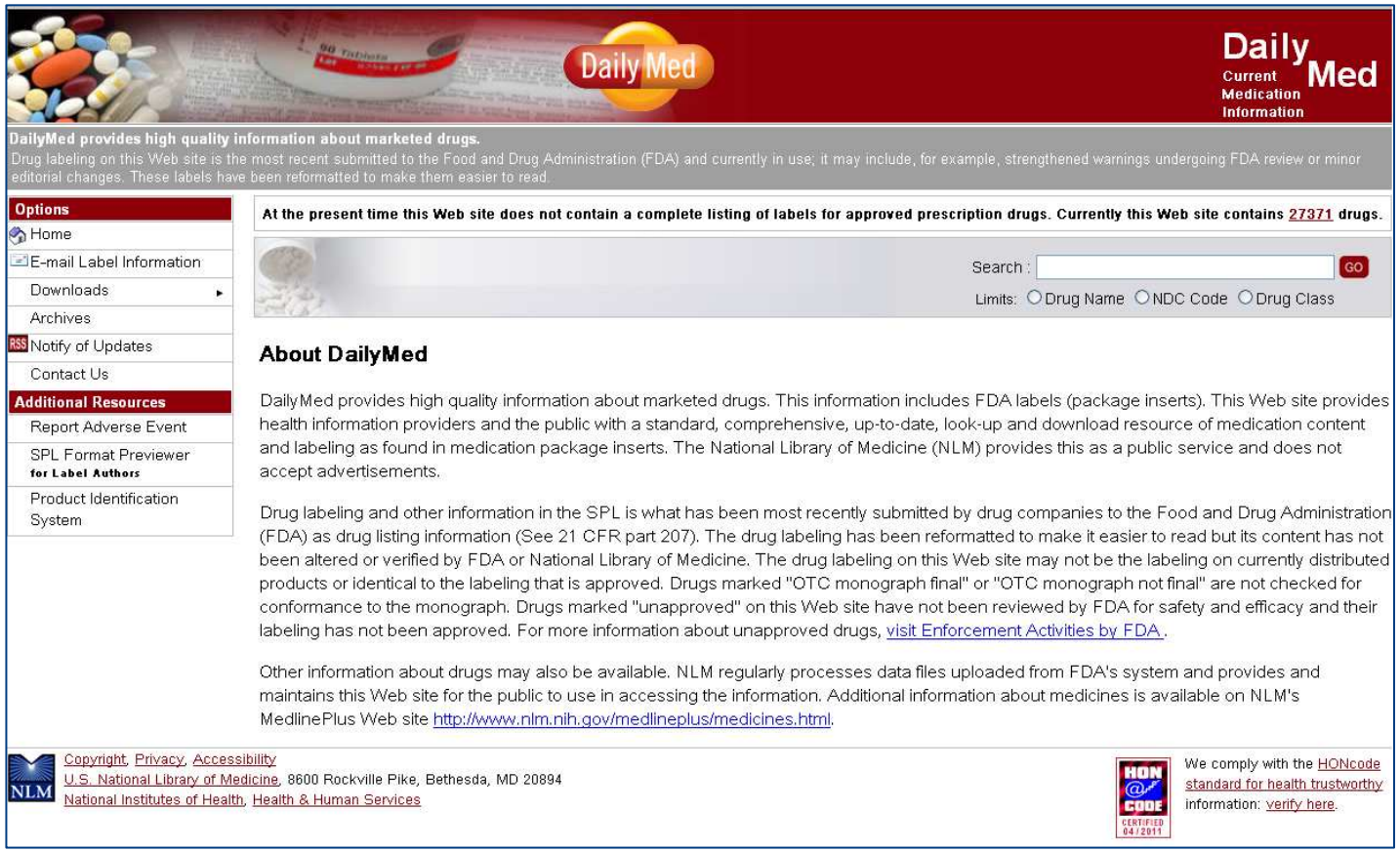
BioWisdom © 2011

Background and Aims

- DailyMed: online resource of drug package inserts
- Metadata around section names, marketing information, etc.
  - Query by Name, NDC Code, Drug Class
- Data within each section is unstructured
- No mark up of important concepts: indications, adverse events, etc.
- Impossible to ask simple questions
  - Which drugs treat asthma
  - Do NSAIDs induce dyspnoea?
- Repurpose labels to create actionable intelligence
- Metawise used to mark up key language from SNOMED CT
  - Drugs, AEs, boxed warnings, lab tests
- Translation of free text descriptors to SNOMED CT
  - Creates metadata index of SNOMED terms across the entire database
- Manual curation of a subset to create *assertional* metadata

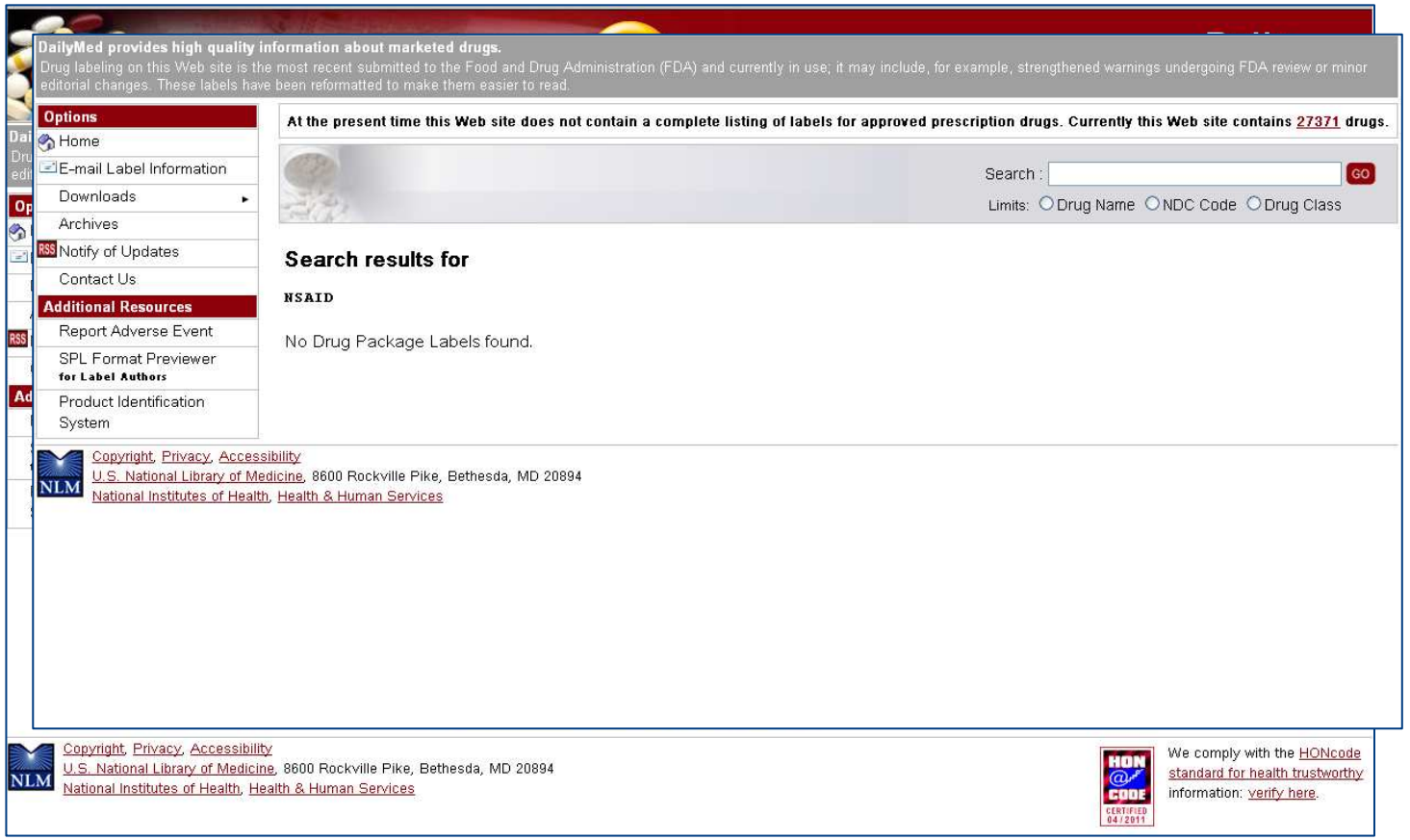
BioWisdom © 2011

DailyMed



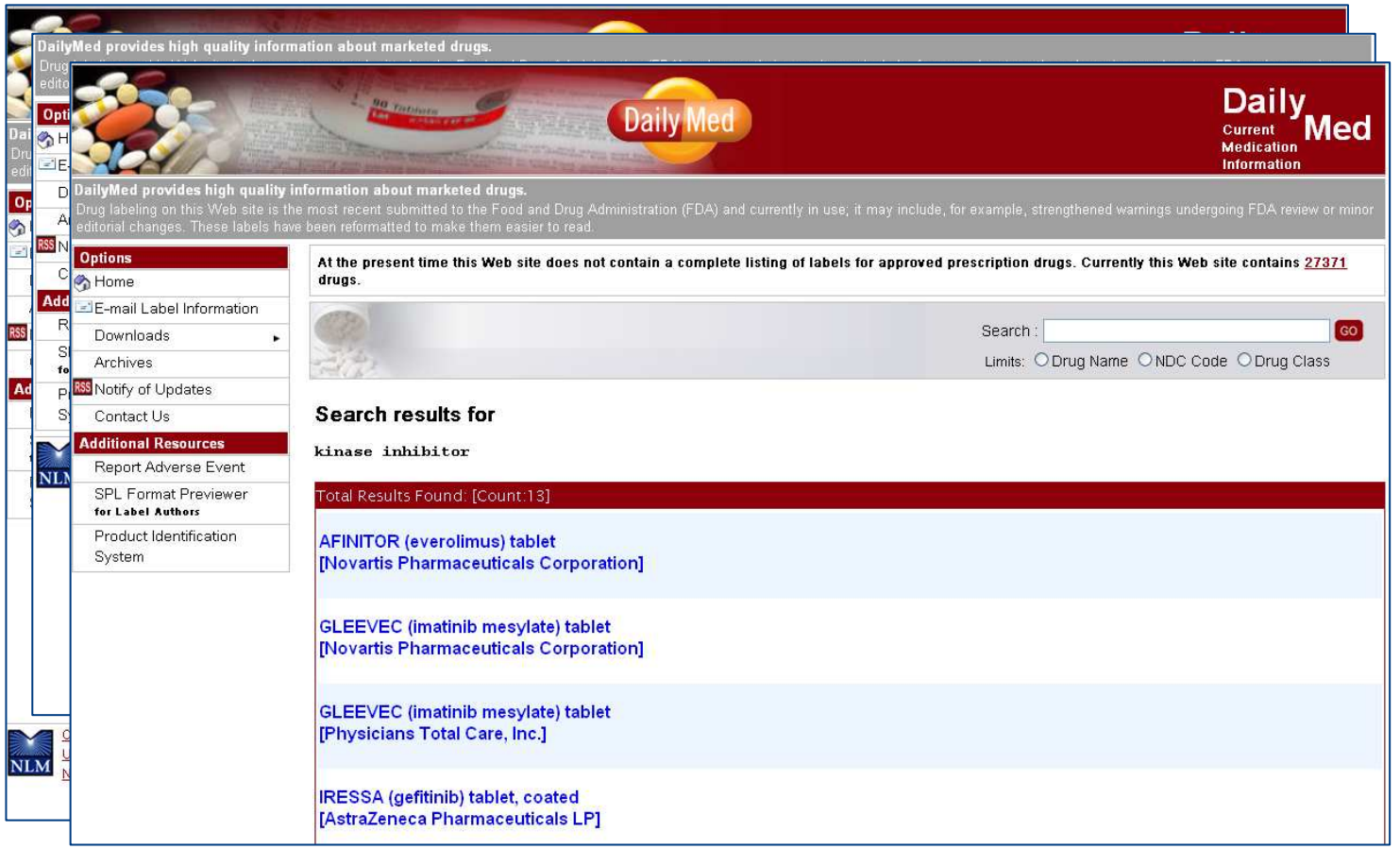
BioWisdom © 2011

DailyMed



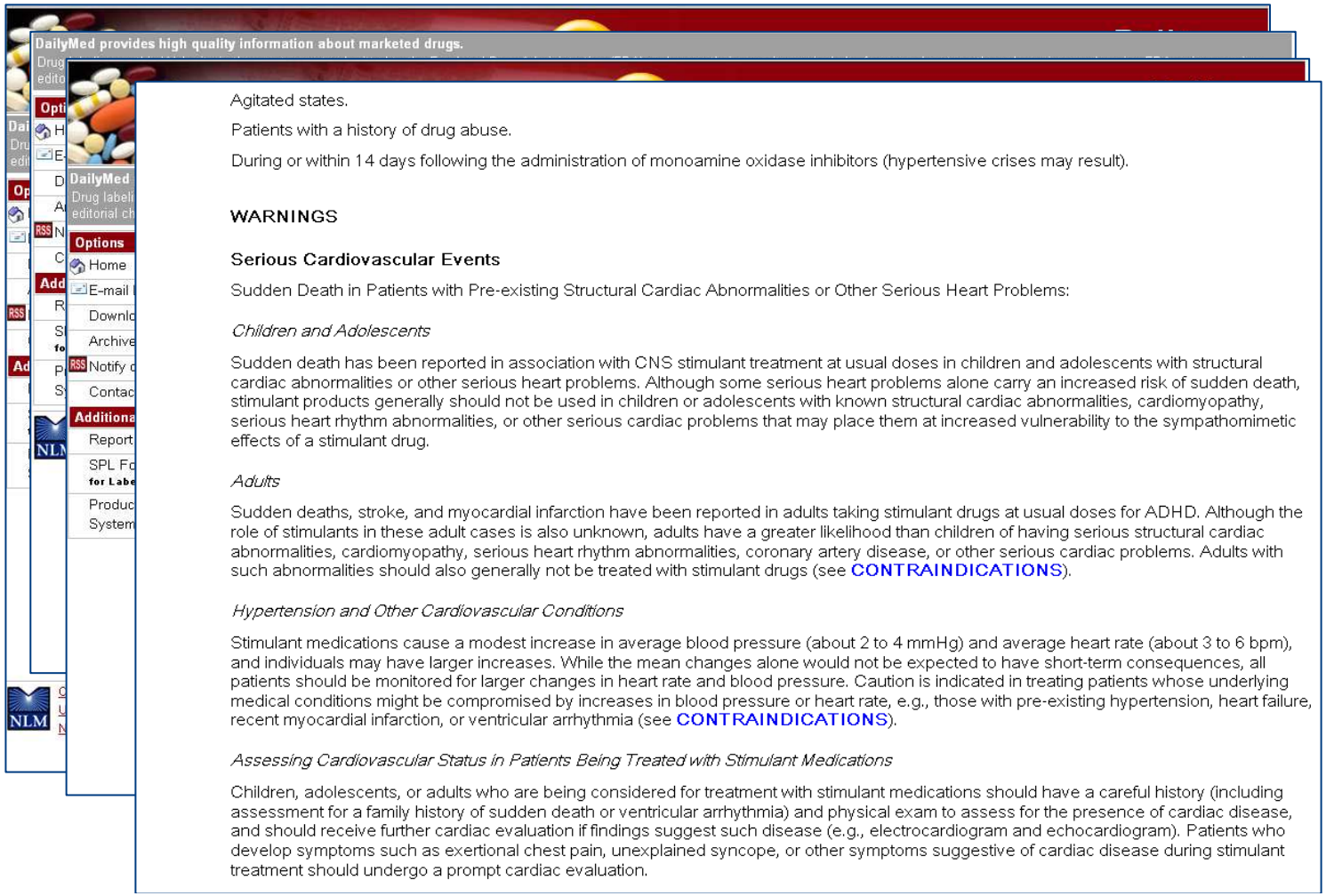
BioWisdom © 2011

DailyMed



BioWisdom © 2011

DailyMed

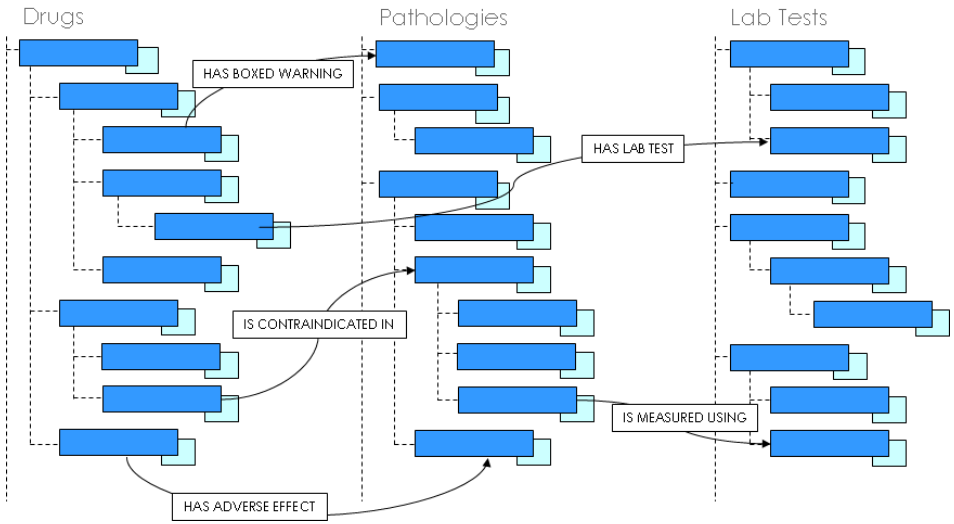


© 2011

Assertional Metadata

- Combines concepts as subject-relationship-object triplet, e.g.

Amiodarone – PROLONGS – QT Interval

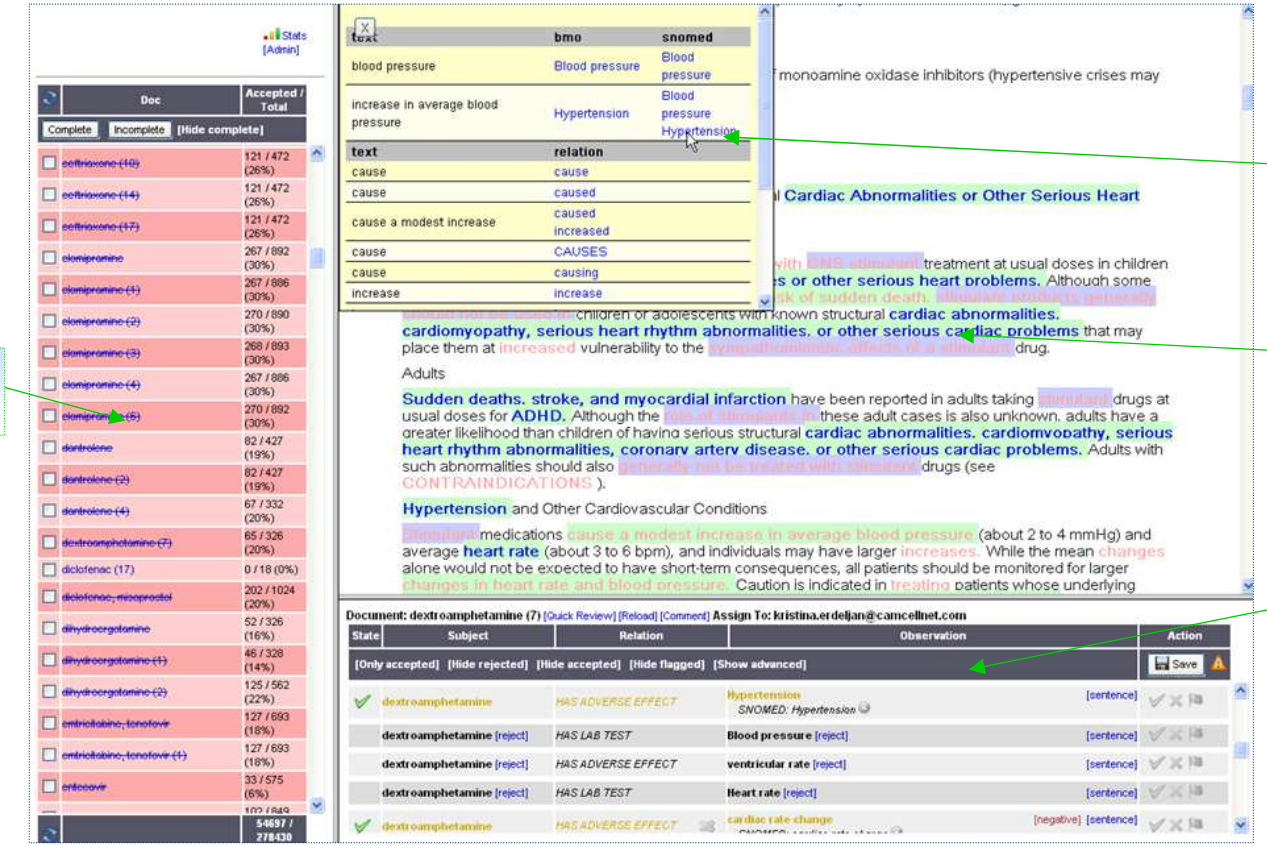


- Adds extra layer of detail over simple tagging of key terms
- Creates a navigable information layer over documents
  - Relationships add context
  - Is myalgia an indication, adverse effect or boxed warning?
- Allows more complex questions to be posed
  - What adverse effects are observed in drugs that treat breast cancer?
  - Which black box warnings occur in medicines that list haemolytic anaemia as an adverse event?
- When combined with taxonomies, can ask broader, class related questions
  - What testing should be carried out if prescribing a *corticosteroid*?
  - Which *LFTs* should be performed with *statins*?

BioWisdom © 2011

Mark-up and Curation of Drug Labels

- Bespoke curation interface
- Web based, allowing remote users to access concurrently



BioWisdom © 2011

Entity Recognition from Drug Labels

SNOMED CT Concept	DailyMed Match
Anaphylactic-type reaction	anaphylactoid reaction, anaphylactic reactions, anaphylactic shock, anaphylactoid and allergic reactions, anaphylactic reaction
Hypertension	elevation of blood pressure, increase in average blood pressure, high blood pressure, increased blood pressure

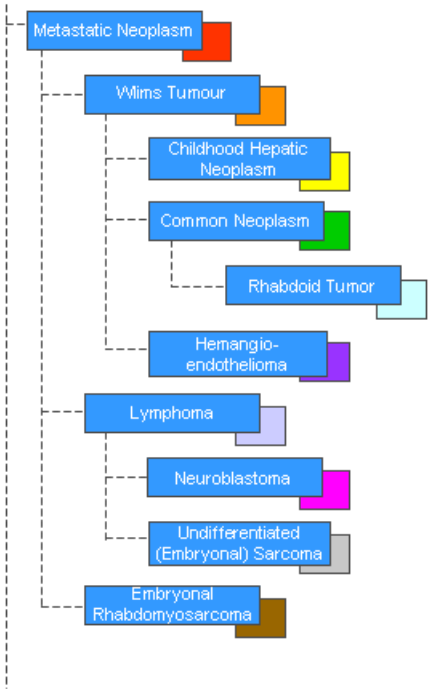
- Stemming
  - Analphylactic/anaphylactoid, reaction/reactions
- Dictionary Look-ups
  - Hypertension -> increased blood pressure
- Stemming, word insertion
  - increased blood pressure -> increased **in average** blood pressure

BioWisdom © 2011

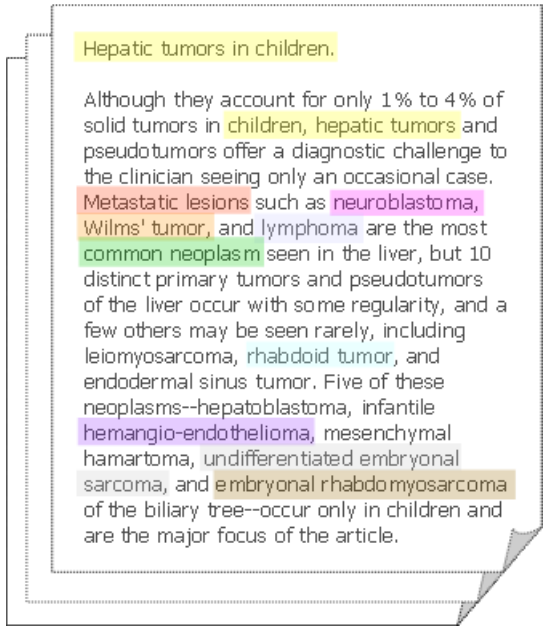
The Benefits of Organised Data

- Document mark-up and translation creates metadata index of SNOMED terms across the entire database
  - Enables semantic querying, document triage
  - Adopt any target ontology structure for taxonomic querying

Target List, e.g. NCI Thesaurus Neoplasms



Document Source, e.g. Medline



Query at most granular level for individual concepts

BioWisdom © 2011

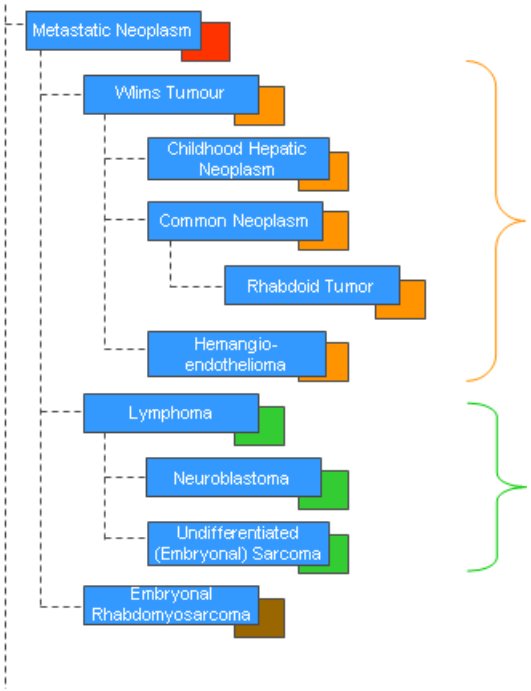
The Benefits of Organised Data

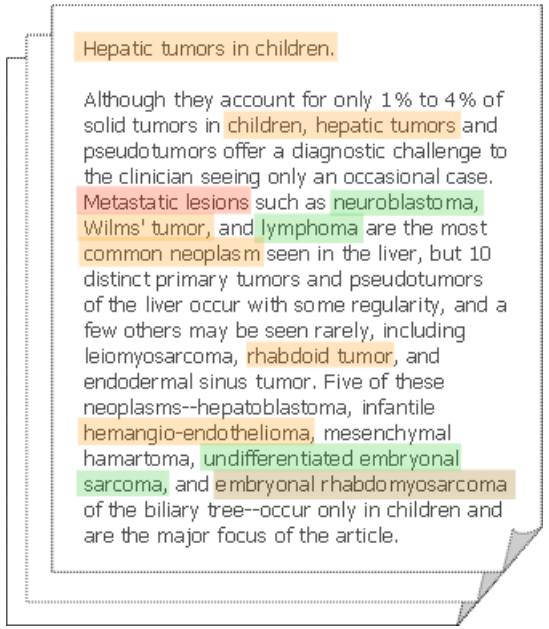
- Document mark-up and translation creates metadata index of SNOMED terms across the entire database
  - Enables semantic querying, document triage
  - Adopt any target ontology structure for taxonomic querying

Target List, e.g. NCI Thesaurus Neoplasms

Document Source, e.g. Medline

Query at selected parent level for groups of related concepts





BioWisdom © 2011

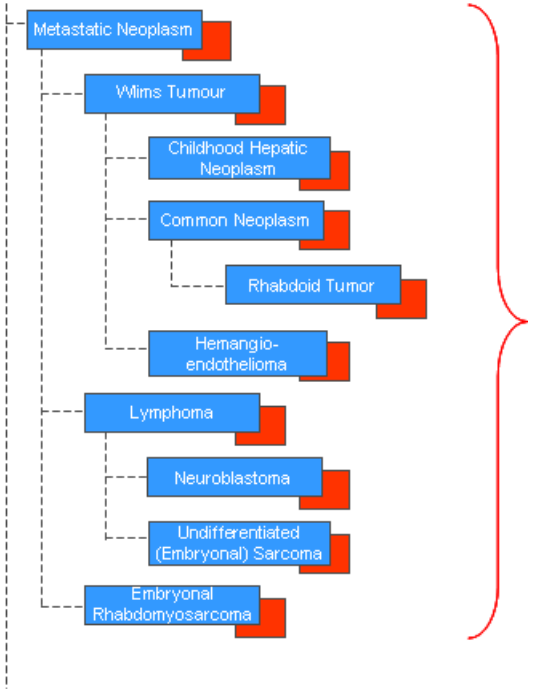
The Benefits of Organised Data

- Document mark-up and translation creates metadata index of SNOMED terms across the entire database
  - Enables semantic querying, document triage
  - Adopt any target ontology structure for taxonomic querying

Target List, e.g. NCI Thesaurus Neoplasms

Document Source, e.g. Medline

Query at selected top level i.e. Metastatic Neoplasm



Hepatic tumors in children.

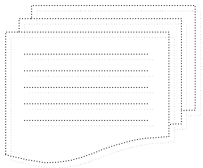
Although they account for only 1% to 4% of solid tumors in children, hepatic tumors and pseudotumors offer a diagnostic challenge to the clinician seeing only an occasional case. Metastatic lesions such as neuroblastoma, Wilms' tumor, and lymphoma are the most common neoplasm seen in the liver, but 10 distinct primary tumors and pseudotumors of the liver occur with some regularity, and a few others may be seen rarely, including leiomyosarcoma, rhabdoid tumor, and endodermal sinus tumor. Five of these neoplasms--hepatoblastoma, infantile hemangio-endothelioma, mesenchymal hamartoma, undifferentiated embryonal sarcoma, and embryonal rhabdomyosarcoma of the biliary tree--occur only in children and are the major focus of the article.

BioWisdom © 2011

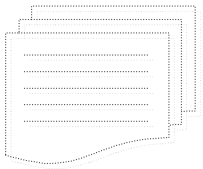
Interoperability with Other Resources

- Several reports of "over-warning" in drug labelling
  - Duke *et al*, "Archives of Internal Medicine", July 2011
- FDA took action in 2006 to prevent to discourage this act
- Compare DailyMed boxed warnings with reported effects in AERS
- Identify similarities / gaps
- DailyMed and AERS expressed consistently in BioWisdom's Safety Intelligence Program (SIP)

DailyMed (SNOMED CT)



AERS (MedDRA)



Drug	DailyMed Warning	In AERS?
Bleomycin	Confusion	✓
	Fever	✓
	Hypotension	✓
	Hypothermia	✓
	Pneumonitis	✓
	Pulmonary fibrosis	✓
	Pulmonary toxicity	✓
	Wheezing	✗
Carbamazepine	Agranulocytosis	✓
	Aplastic anemia	✓
	Bone marrow suppression	✓
	Fever	✓
	Leucopenia	✓
	Low platelet count	✗
	Stevens-johnson syndrome	✓
	Toxic epidermal necrolysis	✓

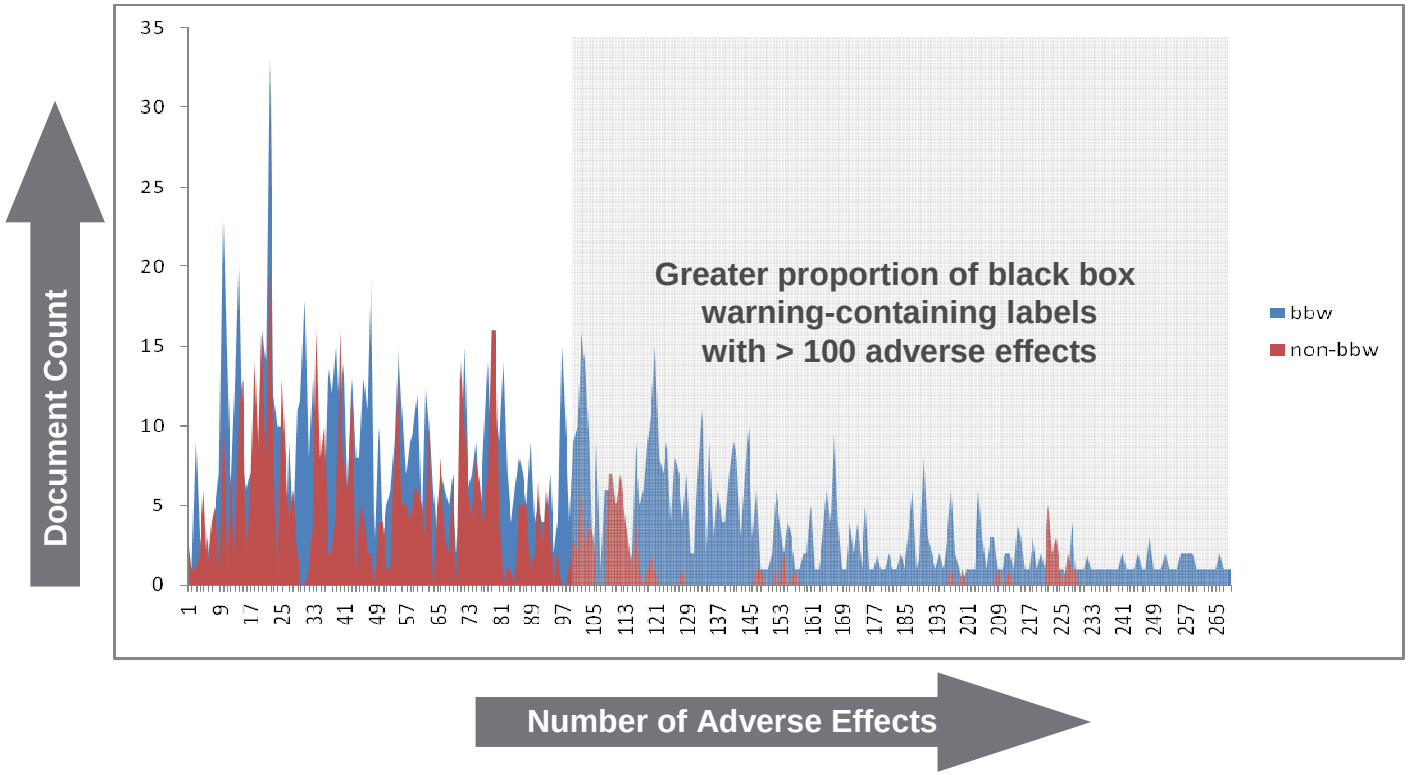
BioWisdom © 2011

Confirmation of Labelled Adverse Effects/Boxed Warnings by AERS

Adverse Events		Boxed Warnings
	Strongly Confirmed	
Macrocytic anemia		Suicidality
Acute myelocytic leukemia		Steatosis
Torsade De Pointes		Adult maltreatment syndrome
Stevens-Johnson syndrome		Myocardial infarction
Long QTc Interval		Abnormal liver function
Stomach cramping		Bone marrow suppression
Toxic nephrosis		Thrombocytopenia
	Poorly Confirmed	

BioWisdom © 2011

Adverse Events per Label:
Comparison of Black Box and Non-Black Box Drugs



BioWisdom © 2011

Consistent Language Facilitates Alignment

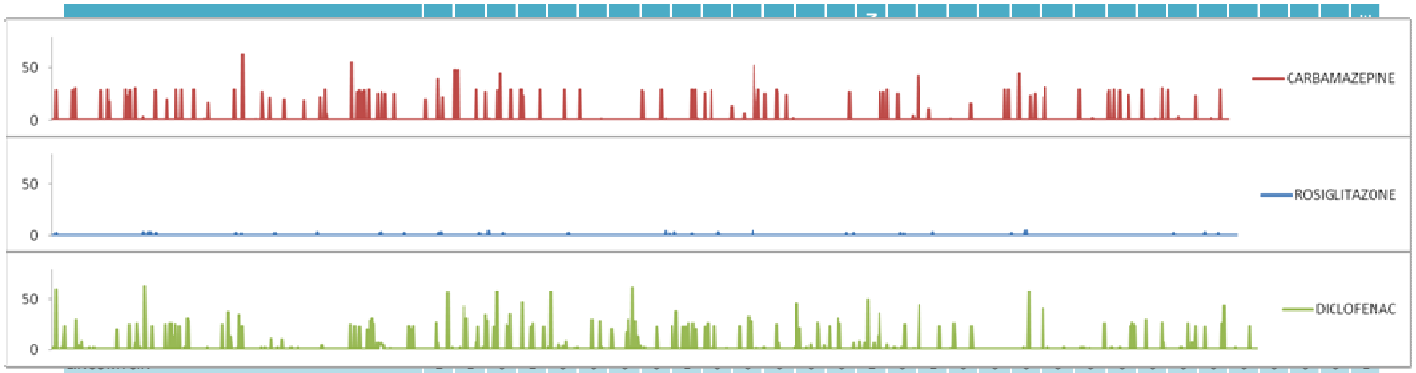
- Align features and record frequency across all labels
- Creates characteristic profile for each label, or each drug
- Tractable to alignment algorithms
  - Cluster drugs according to adverse event profile
  - Could similar profiles be predictive for potential black box warnings in drugs not carrying a warning?
  - Could some profiles act as early warning signals – how does my drug compare to withdrawn compounds?

Drug Name	NAUSEA	EMESIS	HEADACHE	DIARRHEA	RASH	DIZZINESS	INSOMNIA	THROMBOCYTOPENIA	URTICARIA	FEVER	HYPOTENSION	DROWSINESS	PRURITUS	CONSTIPATION	ANAPHYLACTIC-TYPE REACTION	ABDOMINAL PAIN	LEUCOPENIA	CONFUSION	EDEMA	DYSPNEA	ANOREXIA	ANXIETY	SYNCOPE	XEROSTOMIA	ARTHRALGIA	TACHYCARDIA	PARESTHESIA	DEATH	MYALGIA	ALOPECIA	STEVENS-JOHNSON SYNDROME
COLISTIMETHATE	0	0	0	0	1	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
AMLODIPINE & TELMISARTAN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0
ERGOTAMINE & CAFFEINE	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0
OXYTOCIN	2	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PERINDOPRIL	3	3	3	3	3	3	0	3	0	3	3	3	3	3	5	5	0	0	3	3	0	3	3	3	3	0	3	3	3	0	0
NOREPINEPHRINE	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
LINCOMYCIN	1	1	0	1	0	0	0	0	1	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
PERFLUTREN	1	1	1	1	0	1	0	0	2	1	2	0	2	0	1	1	1	0	0	1	0	0	1	1	1	1	1	2	0	0	0
PIOGLITAZONE & METFORMIN	0	0	2	2	0	0	0	0	0	0	1	1	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	3	0	0
IFOSFAMIDE	2	2	0	2	0	2	0	2	0	0	0	4	0	2	0	0	2	6	0	0	2	0	0	0	0	0	0	0	0	3	0
BUDESONIDE	5	2	8	2	9	2	3	0	10	6	1	2	5	0	6	1	0	2	0	2	4	5	1	1	3	2	2	5	7	2	0
LISINOPRIL	66	37	0	0	32	0	33	34	33	59	63	33	34	34	48	61	38	33	35	34	34	0	53	34	67	0	33	55	66	33	33
METHOCARBAMOL	17	17	17	0	17	17	17	0	17	17	17	17	17	0	17	0	17	17	4	0	0	0	17	0	0	0	0	0	0	0	0
ALLOPURINOL	13	13	12	12	7	1	1	13	9	14	0	22	18	0	0	13	22	1	0	0	13	0	0	0	20	0	12	15	2	13	20

BioWisdom © 2011

Consistent Language Facilitates Alignment

- Align features and record frequency across all labels
- Creates characteristic profile for each label, or each drug
- Tractable to alignment algorithms
  - Cluster drugs according to adverse event profile
  - Could similar profiles be predictive for potential black box warnings in drugs not carrying a warning?
  - Could some profiles act as early warning signals – how does my drug compare to withdrawn compounds?

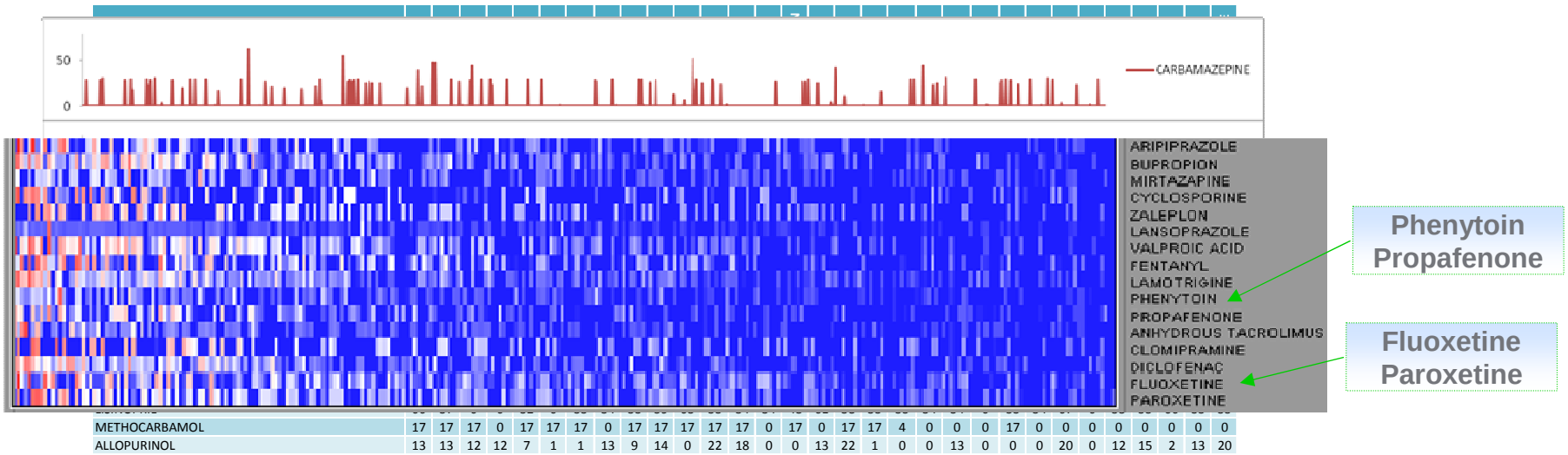


PERFLUTREN	1	1	1	1	0	1	0	0	2	1	2	0	2	0	1	1	1	0	0	1	0	0	1	1	1	1	2	0	0	0
PIOGLITAZONE & METFORMIN	0	0	2	2	0	0	0	0	0	0	1	1	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	3	0	0
IFOSFAMIDE	2	2	0	2	0	2	0	2	0	0	0	4	0	2	0	0	2	6	0	0	2	0	0	0	0	0	0	0	3	0
BUDESONIDE	5	2	8	2	9	2	3	0	10	6	1	2	5	0	6	1	0	2	0	2	4	5	1	1	3	2	2	5	7	2	0
LISINOPRIL	66	37	0	0	32	0	33	34	33	59	63	33	34	34	48	61	38	33	35	34	34	0	53	34	67	0	33	55	66	33	33
METHOCARBAMOL	17	17	17	0	17	17	17	0	17	17	17	17	17	0	17	0	17	17	4	0	0	0	17	0	0	0	0	0	0	0
ALLOPURINOL	13	13	12	12	7	1	1	13	9	14	0	22	18	0	0	13	22	1	0	0	13	0	0	0	20	0	12	15	2	13	20

BioWisdom © 2011

Consistent Language Facilitates Alignment

- Align features and record frequency across all labels
- Creates characteristic profile for each label, or each drug
- Tractable to alignment algorithms
  - Cluster drugs according to adverse event profile
  - Could similar profiles be predictive for potential black box warnings in drugs not carrying a warning?
  - Could some profiles act as early warning signals – how does my drug compare to withdrawn compounds?

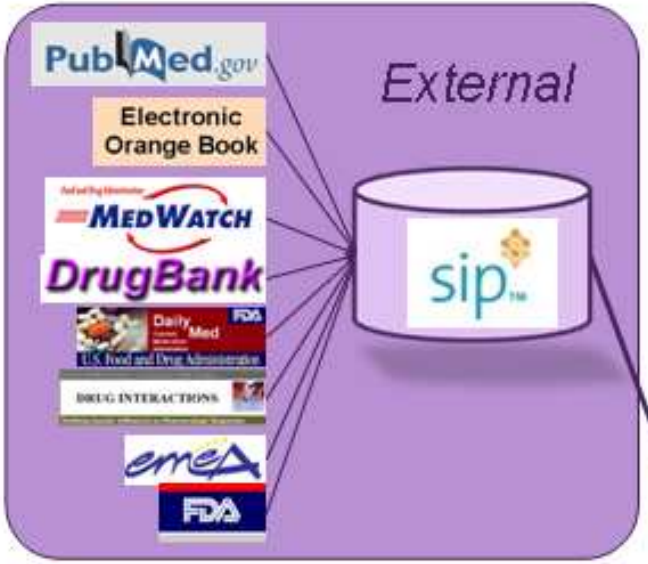


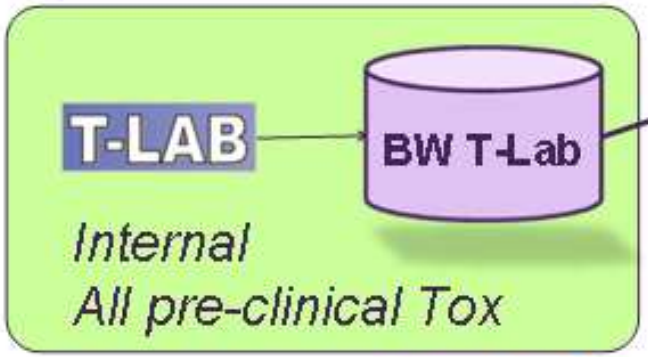
BioWisdom © 2011

Safety Intelligence: Systems Toxicology

Vision

Augmentation of drug safety decision-making through integration of internal and external safety prior knowledge

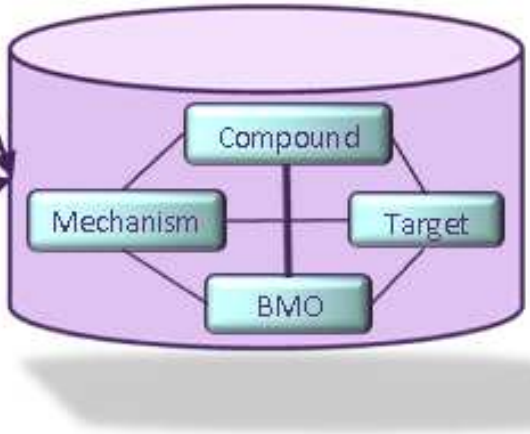




Partnership with **Biowisdom**

Typical Business Questions Addressed

- What compounds/targets have be tested preclinically and what are the findings ?
- What compounds/targets show preclinical tox in 28 day rat studies?
- "Find all compounds causing glycogen accumulation in liver in man and rat" & "What are their common pharmacology and pathways?"



Integrated views in PharmaConnect

Conclusions

- DailyMed is a rich source of drug intelligence
- Unstructured, suffers all of the drawbacks typical of free text
- Hinders its use in decision support
  - Querying, analysis, integration with other applications
- Concept mark-up and translation with Metawise
- Curation of assertional metadata
  - Actionable information layer
- Metadata index provides powerful substrate for analysis
  - Cluster drugs or individual labels
  - Integrate with other data sources
- Conversion of data to standards is of increasing importance
  - SEND from C-DISC, interchange between organisations and regulators

BioWisdom © 2011

biowisdom
Part of the Instem Group

Contact:
paul.bradley@biowisdom.com
paul.bradley@instem.com

BioWisdom © 2011